

SEQUENCE FROM N.A.
MEDLINE=9167334; PubMed=10066439;
Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.,
"Lipophilins: human peptides homologous to rat prostatein.,
Biochem. Biophys. Res. Commun. 756:147-155(1999).
[3]
SEQUENCE OF 19-85.
TISSUE=Testis;
MEDLINE=96163342; PubMed=9504814;
Molloy M.P., Bollis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,
Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.,
"Establishment of the human reflex tear two-dimensional polyacrylamide
gel electrophoresis reference map: new proteins of potential
diagnostic value.,
Electrophoresis 18:2811-2815(1997).
[4]
SEQUENCE OF 19-46 AND 60-78, AND MASS SPECTROMETRY.
MEDLINE=9835871; PubMed=9720917;
Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Qu X.-D., Martin D.,
Glasgow B.J.,
"Lipophilin, a novel heterodimeric protein of human tears.,
FEBS Lett. 432:163-167(1998).
-!- FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS, MAY ALSO BIND
ESTRADIOL, A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.
-!- MAY BE UNDER TRANSCRIPTIONAL REGULATION OF STEROID HORMONES.
-!- SUBUNIT: HETERODIMER OF A LIPOPHILIN A AND A LIPOPHILIN B
(MAMMAGLOBIN B) MONOMER ASSOCIATED HEAD TO HEAD.
-!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRACHEA, KIDNEY, STEROID
RESPONSIVE TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY),
AND SALIVARY GLAND.
-!- MASS SPECTROMETRY: MW=8854.94; METHOD=ELECTROSPRAY; RANGE=19-95.
-!- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF071219; AAC79996.1; -;
DR EMBL; AJ224173; CAA11865.1; -;
DR MIM; 604398; -;
DR InterPro; IPR000329; -;
DR Pfam; PF01099; Uteroglobin; 1;
DR PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE; PS00404; UTEROGLOBIN_2; FALSE_NEG.
KW Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 95 MAMMAGLOBIN B.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 95 AA; 10884 MW; 0719738289P8F8D CRC64;

Query Match 99.5%; Score 390; DB 1; Length 95;
Best Local Similarity 98.6%; Pred. No. 1.4e-35;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFG 60
19 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFG 78
61 LMMHTVYDSIWCNL 74
19 LMMHTVYDSIWCNM 92

RESULT 2
MGBA_HUMAN STANDARD; PRT; 93 AA.
Q13296;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
MAMMAGLOBIN A PRECURSOR (MAMMAGLOBIN 1).
MGBl OR UGB2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Breast;
MEDLINE=96223698; PubMed=8631025;
Watson M.A., Fleming T.P.;
"Mammaglobin, a mammary-specific member of the uteroglobin gene family is overexpressed in human breast cancer.";
Cancer Res. 56:860-865(1996).
[2]
SEQUENCE FROM N.A.
MEDLINE=98147371; PubMed=9488047;
Watson M.A., Darlow C., Zimonjic D.B., Popescu N.C., Fleming T.;
"Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13.";
Oncogene 16:817-824(1998).
-1- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST CANCER.
-1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U33147; AAC50375.1; -;
DR EMBL; AF015224; AAC39608.1; -;
DR MIM; 605562; -;

DR InterPro; IPR000329; -;
DR Pfam; PF01099; Uteroglobin; 1;
DR PROSITE; PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE; PS00404; UTEROGLOBIN_2; 1;
KW Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 93 MAMMAGLOBIN A.
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 93 AA; 10499 MW; 2896E8C43BF053E2 CRC64;

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 5.9e-16;
Matches 39; Conservative 12; Mismatches 42; Indels 0; Gaps 0;

QY 2 SGCILLEDMVKTINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFL 61
19 SGCILLEDMVKTINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNKFL 79

QY 62 LMMHTVYDSIWCNL 74
19 LMMHTVYDSIWCNM 92

RESULT 3

PSC3_RAT
ID PSC3_RAT STANDARD; PRT; 95 AA.
AC P02780; O63463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE PROSTATIC STEROID-BINDING PROTEIN C3 CHAIN PRECURSOR (PROSTATEIN
DE PEPTIDE C3).
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=83082848; PubMed=6294095;
Parker M.G., White R., Hurst H., Needham M., Tilly R.;
"Prostatic steroid-binding protein. Isolation and characterization of
C3 genes.";
J. Biol. Chem. 258:12-15(1983).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=83238526; PubMed=6190812;
Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M.,
French F.S.;
"Isolation of two genomic sequences encoding the Mr = 14,000 subunit
of rat prostatein.";
J. Biol. Chem. 258:8861-8866(1983).
[3]
SEQUENCE FROM N.A.
RX MEDLINE=92165796; PubMed=1537831;
Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
French F.S.;
"Response elements of the androgen-regulated C3 gene.";
J. Biol. Chem. 267:4456-4466(1992).
[4]
ERRATUM.
RX MEDLINE=92218467; PubMed=1339454;
Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
French F.S.;
J. Biol. Chem. 267:7958-7958(1992).

Copied from PCT009367009 on 27-07-2004

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:13:38 ; Search time 10.24 seconds
(without alignments)
247.549 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGCKLEDVMTKINSDIS.....TLNFGMLMHTVDSIWCNL 74

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries
Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	99.5	95	1	MGBB_HUMAN
2	208	53.1	93	1	MGSA_HUMAN
3	163	41.6	95	1	PSG3_RAT
4	70	17.9	303	1	FTSY_RICPR
5	66.5	17.0	234	1	RADC_HAEIN
6	65.5	16.7	131	1	YR59_MYCTU
7	65.5	16.7	603	1	VEI_HPV08
8	64.5	16.5	603	1	VEI_HPV21
9	64.5	16.5	611	1	DYNL_ARATH
10	61.5	15.7	162	1	IL15_MOUSE
11	61.5	15.7	570	1	HEML_KUULA
12	61.5	15.7	734	1	METE_THEMA
13	61	15.6	605	1	VEI_HPV14
14	61	15.6	663	1	TA21_SCHPO
15	61	15.6	732	1	KU86_MOUSE
16	60.5	15.4	430	1	SERC_ARATH
17	60.5	15.4	684	1	TC10_YEAST
18	60	15.3	268	1	TRCP_ACICA
19	60	15.3	606	1	VEI_HPV5B
20	60	15.3	790	1	V90K_AMYLE
21	60	15.3	1071	1	PR16_YEAST
22	59.5	15.2	162	1	IL15_CERAE
23	59.5	15.2	162	1	IL15_HUMAN
24	59.5	15.2	162	1	IL15_MACMU
25	59.5	15.2	925	1	YESG_SCHPO
26	59.5	15.2	1177	1	X307_MYCGE
27	59	15.1	604	1	VEI_HPV12
28	59	15.1	606	1	VEI_HPV05
29	59	15.1	690	1	Y023_NPVAC
30	59	15.1	2144	1	GLT1_YEAST
31	59	15.1	4540	1	DYHC_PASTE
32	58.5	14.9	322	1	DIDH_RAT
33	58.5	14.9	606	1	RA17_SCHPO

34	58	14.8	113	1	PI13_MYCMV	003290 mycoplasma
35	58	14.8	1128	1	DNBI_HSVSA	P24910 herpesvirus
36	57.5	14.7	162	1	IL15_PIG	Q95253 sus scrofa
37	57.5	14.7	300	1	MOV_P_AMYLE	P03595 alfalfa mos
38	57.5	14.7	359	1	PD12_SCHPO	O13811 schizosacch
39	57.5	14.7	527	1	G19P_HUMAN	P14314 homo sapien
40	57.5	14.7	756	1	METE_BACHD	O9kfl1 bacillus ha
41	57.5	14.7	871	1	SG10_YEAST	O06245 saccharomyc
42	57	14.5	102	1	BOLA_VIBAL	O05585 vibrio algi
43	57	14.5	281	1	E2F6_HUMAN	O73461 homo sapien
44	57	14.5	299	1	HIS1_ECOLI	P10366 escherichia
45	57	14.5	604	1	VE1_HPV36	P50808 human papil

ALIGNMENTS

RESULT	1
MGBB_HUMAN	
ID	MGBB_HUMAN STANDARD; PRT; 95 AA.
AC	O75556;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	MAMMAGLOBIN B PRECURSOR (MAMMAGLOBIN 2) (LIPOPHILIN C) (LACRYGLOBIN).
GN	MG2 OR UGB3 OR LIPHC.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99026127; PubMed=9806831;
RA	Becker R.M., Darrow C., Zimonjic D.B., Popescu N.C., Watson M.A.,
RA	Fleming T.P.;
RT	"Identification of mammaglobin B, a novel member of the uteroglobin
RT	gene family".
RL	Genomics 54:70-78(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99167354; PubMed=10066439;
RA	Zhao C., Nguyen T., Yusifov T., Glasgow B.J., Lehrer R.I.;
RA	"Lipophilins: human peptides homologous to rat prostatein.";
RL	Biochem. Biophys. Res. Commun. 256:147-155(1999).
RN	[3]
RP	SEQUENCE OF 19-85.
RC	TISSUE=Tears;
RC	MEDLINE=98163342; PubMed=9504814;
RA	Wolloy M.P., Bollis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,
RA	Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;
RT	"Establishment of the human reflex tear two-dimensional polyacrylamide
RT	gel electrophoresis reference map: new proteins of potential
RL	diagnostic value.";
RL	Electrophoresis 18:2811-2815(1997).
RN	[4]
RP	SEQUENCE OF 19-46 AND 60-78, AND MASS SPECTROMETRY.
RA	MEDLINE=98385871; PubMed=9720917;
RA	Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Ou X.-D., Martin D.,
RA	Glasgow B.J.;
RT	"Lipophilin, a novel heterodimeric protein of human tears.";
RL	FEBS Lett. 432:163-167(1998).
CC	-!- FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS, MAY ALSO BIND
CC	ESTRADIOL, A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.
CC	MAY BE UNDER TRANSCRIPTIONAL REGULATION OF STEROID HORMONES.
CC	-!- SUBUNIT: HETEROLOGOUS OF A LIPOPHILIN A AND A LIPOPHILIN C
CC	(MAMMAGLOBIN B) MONOMER ASSOCIATED HEAD TO HEAD.
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRACHEA, KIDNEY, STERIOD
CC	RESPONSIVE TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY),
CC	AND SALIVARY GLAND.
CC	-!- MASS SPECTROMETRY: MW=8854.94; METHOD=ELECTROSPRAY; RANGE=19-95.
CC	-!- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC	SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF071219; AAC79996.1; -
EMBL: AJ224173; CAAL1865.1; -
MIM: 604396; -
Pfam: PF01099; Uteroglobin; 1.
PROSITE: PS00403; UTEROGLBIN_1; FALSE_NEG.
PROSITE: PS00404; UTEROGLBIN_2; FALSE_NEG.
Signal: Glycoprotein.
SIGNAL 1 18
CHAIN 19 95
CARBOHYD 68 68
SEQUENCE 95 AA; 10884 MW; 0719738289F8F8D CRC64;
MAMMAGLOBIN B.
N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 99.5%; Score 390; DB 1; Length 95;
Best Local Similarity 98.6%; Pred. No. 1.4e-35;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DSGCKLEDVMTKINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNFG 60
|||||
19 DSGCKLEDVMTKINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNFG 78

61 LMHHTVYDSIWCNL 74

|||||
79 LMHHTVYDSIWCNM 92

RESULT 2
ID MGBA_HUMAN STANDARD; PRT; 93 AA.
Q13296;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
MAMMAGLOBIN A PRECURSOR (MAMMAGLOBIN 1).
MGB1 OR UGB2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE-Breast;
MEDLINE=96223698; PubMed=8631025;
Watson M.A., Fleming T.P.;
"Mamaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer.";
Cancer Res. 56:860-865(1996).
[2]
SEQUENCE FROM N.A.
MEDLINE=98147371; PubMed=9488047;
Watson M.A., Darrow C., Zimonjic D.B., Popescu N.C., Fleming T.;
"Structure and transcriptional regulation of the human mamaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13.";
Oncogene 16:817-824(1998).
-1- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST CANCER.

-1- SIMILARITY: BELONGS TO THE UTEROGLBIN FAMILY. LIPOPHILIN SUBFAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: U33147; AAC50375.1; -
EMBL: AF015224; AAC39608.1; -
MIM: 605562; -
InterPro: IPR000329; -
Pfam: PF01099; Uteroglobin; 1.
PROSITE: PS00403; UTEROGLBIN_1; FALSE_NEG.
PROSITE: PS00404; UTEROGLBIN_2; 1.
Signal: Glycoprotein.
SIGNAL 1 18
CHAIN 19 93
CARBOHYD 53 53
FT CARBOHYD 68 68
FT CARBOHYD 68 68
SQ SEQUENCE 93 AA; 10499 MW; 2896E8C43BF053E2 CRC64;

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 5.9e-16;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 SGCKLEDVMTKINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNFG 61
|||||
20 SGCKLEDVMTKINSISIPYKELQEFIDSDAAAEAMGKFCFLNQSHRTLNFG 79

QY 62 MMHVTYDSIWCNL 74

|||||
80 FMQLIVDSIWCNL 92

RESULT 3
ID PSC3_RAT STANDARD; PRT; 95 AA.
P02780; Q63463;
21-JUL-1986 (Rel. 01, Created)
21-JUL-1986 (Rel. 01, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
DE PROSTATIC STEROID-BINDING PROTEIN C3 CHAIN PRECURSOR (PROSTATEIN PEPTIDE C3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
MEDLINE=83082848; PubMed=6294095;
Parker M.G., White R., Hurst H., Needham M., Tilly R.;
"Prostatic steroid-binding protein. Isolation and characterization of C3 genes.";
J. Biol. Chem. 258:12-15(1983).
[2]
SEQUENCE FROM N.A.
MEDLINE=83238526; PubMed=6190812;
Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M., French F.S.;
"Isolation of two genomic sequences encoding the Mr = 14,000 subunit of rat prostatein.";
J. Biol. Chem. 258:8861-8866(1983).
[3]
SEQUENCE FROM N.A.
MEDLINE=92165796; PubMed=1537831;
Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M., French F.S.;
"Response elements of the androgen-regulated C3 gene.";
J. Biol. Chem. 267:4456-4466(1992).
[4]
ERRATUM.
MEDLINE=92218467; PubMed=1339454;
Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M., French F.S.;
J. Biol. Chem. 267:7958-7958(1992).

Copied from PCT009367009 on 27-04-2004


```

DR Pfam; PF00519; El; 1.
DR Pfam; PF00524; El; N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
FT Nuclear protein.
NW NP_BIND 431 438 ATP (POTENTIAL).
SQ SEQUENCE 603 AA; 68821 MW; 0813860098D48AD CRC64;

Query Match 16.7%; Score 65.5; DB 1; Length 603;
Best Local Similarity 22.2%; Pred. No. 11;
Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;

QY 1 DSGCKLLEDWVEKINSIDISIP-----EYKELLOEFIDSDAAEAMGK 43
DB 112 DSGVELTLNNEAEVDSVEVPAIDSRDEGGGALDIDYALLR---SSNTRATLMK 168
QY 44 PKQCFLN-----QSHRLKNFGLM---MHTVYDS 69
DB 169 FKEAFGDFGNELTRFKSYKTCNYYVWVAAYAVHDVYES 207

RESULT 8
REL_HPV21 STANDARD; PRT; 603 AA.
AC VE1_HPV21
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN EL.
OS El.
OC Human papillomavirus type 21.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
OX NCBI_TaxID=31548;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H.;
RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
    VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
    PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
    CONTAINS BINDING SITES FOR BOTH PROTEINS.
    -!- SUBCELLULAR LOCATION: NUCLEAR.
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (see http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    -----
    EMBL; U31779; AAA79396.1; -
    InterPro; IPR001177; -
    Pfam; PF00519; El; 1.
    Pfam; PF00524; El; N; 1.
    KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
    NW NP_BIND 431 438 ATP (POTENTIAL).
    SQ SEQUENCE 603 AA; 68983 MW; 93006494BEF10339 CRC64;

Query Match 16.5%; Score 64.5; DB 1; Length 603;
Best Local Similarity 26.3%; Pred. No. 14;
Matches 26; Conservative 15; Mismatches 23; Indels 35; Gaps 6;

QY 1 DSG--CKLLEDWVEKINSIDISIP-----EYKELLOEFIDSDAAEAM 41
DB 112 DSGLECTLTN---EDVSEVEVPALDSQPVAEALQGTVDIHVKELRA---SNNKAILM 165
QY 42 GKFKOCF---LNQSHRTLKNFG-----LMHTVYDSI 70
DB 166 AKFEFFGVGNDLTRQFKSKYTCNANVLSVYAVHDDL 204

```

```

RESULT 9
DYNL_ARATH STANDARD; PRT; 611 AA.
AC P42697;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DYNAMIN-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Hwang I.; Goodman H.M.; Yoon H.W.; Yoon J.-H.; Cho M.J.;
RT "Cloning and characterization of a cDNA clone encoding dynamin-like
    GTP binding protein in Arabidopsis thaliana.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (see http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    -----
    EMBL; L36939; AAB63528.1; -
    InterPro; IPR000375; -
    InterPro; IPR001401; -
    Pfam; PF00350; dynamin; 1.
    Pfam; PF01031; dynamin; 2; 1.
    DR PRINTS; PR00195; DYNAMIN.
    DR PROSITE; PS00410; DYNAMIN; 1.
    KW Motor protein; GTP-binding.
    FT NP_BIND 41 48 GTP (POTENTIAL).
    FT NP_BIND 142 146 GTP (POTENTIAL).
    FT NP_BIND 211 214 GTP (POTENTIAL).
    SQ SEQUENCE 611 AA; 68509 MW; EDE98B79B1BC0D5 CRC64;

Query Match 16.5%; Score 64.5; DB 1; Length 611;
Best Local Similarity 20.7%; Pred. No. 15;
Matches 17; Conservative 21; Mismatches 23; Indels 21; Gaps 3;

QY 6 LLEDVKTINSIDISIPYKELLQEFIDSDAAEAMGKFK-----QC----- 47
DB 432 ILKDLVHKSVNETVELKQYPALRVEV--TNAIESLDKMGSKKATLQVDMCSYLTV 489
QY 48 -FLNQSHRTLKNFGLMHTVYD 68
DB 490 DFFRKLQPDQVEKGNPHTSIFD 511

RESULT 10
IL15_MOUSE STANDARD; PRT; 162 AA.
AC P48346;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INTERLEUKIN-15 PRECURSOR (IL-15).
GN IL15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AB001784; RAD36360.1; -
 CC TIGR; TM1286; -
 CC InterPro: IPR002629; -
 CC Pfam: PF01171; Methionine_synth; 1.
 CC Transferrase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat.
 CC METAL 618 ZINC (BY SIMILARITY).
 CC METAL 620 ZINC (BY SIMILARITY).
 CC METAL 704 ZINC (BY SIMILARITY).
 CC SEQUENCE 734 AA; 85570 MW; 34B19187480D443B CRC64;

Query Match 15.7%; Score 61.5; DB 1; Length 734;
 Best Local Similarity 25.4%; Pred. No. 38;
 Matches 18; Conservative 17; Mismatches 25; Indels 11; Gaps 3;

OY 9 DMVEKINDISIPKELLOEFIDSDAAEAMGK-FKOCFLNQSH-----RTLKNF 59

b 168 NOMEKLELSVSV--YKEVFEKLVENGCKEILVNEPAFVCDLEKAHWDILNLVYRELSEF 225

OY 60 GLMHHTVYDSI 70

b 226 PLTVFTYDSV 236

RESULT 13

VEL HPV14

CC VEL HPV14 STANDARD; PRT; 605 AA.

CC P36721;

CC 01-JUN-1994 (Rel. 29, Created)

CC 01-JUN-1994 (Rel. 29, Last sequence update)

CC 15-JUL-1998 (Rel. 36, Last annotation update)

CC REPLICATION PROTEIN E1.

CC E1

CC Human papillomavirus type 14.

CC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.

CC NCBI_TaxID=10605;

CC SEQUENCE FROM N.A.

CC MEDLINE=94265501; PubMed=8205838;

CC Delliuss H., Hofmann B.;

CC "Primer-directed sequencing of human papillomavirus types.";

CC Curr. Top. Microbiol. Immunol. 186:13-31(1994).

CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF

CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2

CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH

CC CONTAINS BINDING SITES FOR BOTH PROTEINS.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X74467; CA52502.1; -

CC PIR; S36469; S36469.

CC InterPro: IPR001177; -

CC Pfam; PF00519; E1; 1.

CC Pfam; PF00524; E1N; 1.

CC Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;

CC Nuclear protein.

CC NP_BIND 433 440 ATP (POTENTIAL).

CC SEQUENCE 605 AA; 69302 MW; 6702915505AA6491 CRC64;

Query Match 15.6%; Score 61; DB 1; Length 605;

Best Local Similarity 23.7%; Pred. No. 34;

Matches 23; Conservative 16; Mismatches 28; Indels 30; Gaps 4;

OY 1 DSCCKLEDMVEKINTSDISIP-----EYKELLOEFIDSDAAEAMGK 43

b 113 DSGLELTLINEAEDVSEVEPALDSQPVAAOIGIVDIHYITELLRA---SNNKAILMAK 169

OY 44 FKOCF---LNQSHRTLKNFG-----LMHHTVYDSI 70

b 170 FKEAFGVGFNDLTPQFKSYKTCNHNWVLSVYAVHDDL 206

RESULT 14

TAZI_SCHPO

ID TAZI_SCHPO STANDARD; PRT; 663 AA.

AC P79005; O00049;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE TELOMERE LENGTH REGULATOR TAZ1.

GN TAZ1 OR MYB1 OR MYB OR SPAC16A10.07C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OC NCBI_TaxID=4896;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RC MEDLINE=97186578; PubMed=9034194;

RA Cooper J.P., Nimmo E.R., Allshire R.C., Cech T.R.;

RT "Regulation of telomere length and function by a Myb-domain protein

RL in fission yeast."

RL Nature 385:744-747(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: REGULATES TELOMERE LENGTH AND FUNCTION. REQUIRED FOR THE

CC REPRESSION OF TELOMERE-ADJACENT GENE EXPRESSION AND FOR NORMAL

CC MEIOSIS OR SPORULATION. IT MAY BE A NEGATIVE REGULATOR OF THE

CC TELOMERE-REPLICATING ENZYME, TELOMERASE, OR MAY PROTECT AGAINST

CC ACTIVATION OF TELOMERASE-INDEPENDENT PATHWAYS OF TELOMERE

CC ELONGATION. IT MAY BE INVOLVED IN THE INTERACTIONS BETWEEN

CC CHROMOSOMES AND SPINDLE PROTEINS, DISRUPTION OF THESE

CC INTERACTIONS WOULD LEAD TO DEFECTIVE MEIOSIS.

CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).

CC -!- SIMILARITY: BELONGS TO THE MYB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Y09406; CAA70568.1; -

CC EMBL; 297185; CAB10000.1; -

CC InterPro: IPR001005; -

CC PROSITE; PS00037; MYB_1; 1.

CC PROSITE; PS00334; MYB_2; FALSE NEG.

CC PROSITE; PS50090; MYB_3; FALSE NEG.

CC Telomere; Nuclear protein; DNA-binding; Repeat.

FT DNA_BIND 556 612 MYB

SQ SEQUENCE 663 AA; 74647 MW; 5852F21518031152 CRC64;

Query Match

Best Local Similarity

15.6%; Score 61; DB 1; Length 663;

23.1%; Pred. No. 38;

200 GTGATGCCGCTGAGAGGCTATGGGAAATTCAGCAGTGTTTCCTCAAC 249
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
250 CAGTCACATAGAACTCTGAAAACCTTTGGAGCTGATGATGCATACAGTGTA 299
67 rAspSerIleTrpCysAsnLeu 74
|||||
300 CGACAGCATTTGGTGTAAATATG 321

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-263-810-5

seq_documentation_block:
; Sequence 5, Application US/09263810
; Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 476 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-263-810-5

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:

US-09-367-009-3 x US-09-263-810-5

Align seg 1/1 to: US-09-263-810-5 from: 1 to: 476

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsn 17
|||||
100 GATTCGTGGCTGCAAACTCTCGAGGACATGGTTGAAAGACCATCAATC 149
17 rAspIleSerIleProGluTyrlsGluLeuGluInGluPheIleAsps 34
|||||

150 CGACATATATCTACCTGAATACAAAGAGCTTCTTCAAGAGTTTCATAGACA 199
34 exAspAlaIaLaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
200 GTGATGCCCTGCAGAGGCTATGGGAAATTCAGCAGTGTTTCCTCAAC 249
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
250 CAGTCACATAGAACTCTGAAAACCTTTGGAGCTGATGATGCATACAGTGTA 299
67 rAspSerIleTrpCysAsnLeu 74
|||||
300 CGACAGCATTTGGTGTAAATATG 321

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-896-5

seq_documentation_block:
; Sequence 5, Application US/08455896
; Patent No. 5668267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-455-896-5

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:

US-09-367-009-3 x US-08-455-896-5

Align seg 1/1 to: US-08-455-896-5 from: 1 to: 403

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSer 18
|||||
118 TCTGCTGCCCTTATTGGAGATGTGATTTCAGGAATCAATCCACA 167

```

18 pIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAspSerA 35
   :::::||||| ::::::::::::::::::::|
168 AGTGTCTAAGACTGAATACAAAGAACTTCTCAAGAGTTCATAGACGACA 217
   :::::||||| ::::::::::::::::::::|
35 spIalaalaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
   ::::::::::::::::::::|
218 ATCCCACTACAAATGCCATAGATGAATTGAAGAAATGTTTCTTAACCAA 267
   ::::::::::::::::::::|
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
   ::::::::::::::::::::|
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
   ::::::::::::::::::::|
68 pSerIleTIPcysAsnLeu 74
   ||||| ::::::::::|
318 CAGCAGTCCTTTGTGATTTA 336

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-933-149-5

```

seq_documentation_block:
; Sequence 5, Application US/08933149
; Patent No. 5922836
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
; TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,149
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: HENDERSON, MELODIE W.
; REGISTRATION NUMBER: 37,848
; REFERENCE/DOCKET NUMBER: 6029-6040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-933-149-5

```

alignment_scores:
 Quality: 208.00 Length: 73
 Ratio: 3.355 Gaps: 0
 Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
 US-09-367-009-3 x US-08-933-149-5 ..
 Align seg 1/1 to: US-08-933-149-5 from: 1 to: 403

```

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
   ::::::::::|
118 TCRCGCTGCCCTTATTTGGAAATGTGATTCCAGACATCAATCCACA 167
   ::::::::::|
18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
   ::::::::::|
168 AGTGTCTAAGACTGAATACAAAGAACTTCTCAAGAGTTCATAGACGACA 217
   ::::::::::|
35 spIalaalaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
   ::::::::::|
218 ATCCCACTACAAATGCCATAGATGAATTGAAGAAATGTTTCTTAACCAA 267
   ::::::::::|
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
   ::::::::::|
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
   ::::::::::|
68 pSerIleTIPcysAsnLeu 74
   ||||| ::::::::::|
318 CAGCAGTCCTTTGTGATTTA 336

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-09-082-343-5

```

seq_documentation_block:
; Sequence 5, Application US/09082343
; Patent No. 5968754
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,343
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-082-343-5

```

alignment_scores:
 Quality: 208.00 Length: 73
 Ratio: 3.355 Gaps: 0
 Percent Similarity: 84.932 Percent Identity: 53.425

```

alignment_block:
US-09-367-009-3 x US-09-082-343-5
Align seg 1/1 to: US-09-082-343-5 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTTTCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-082-253-5

seq_documentation_block:
; Sequence 5, Application US/09082253
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: NO
; HYPOTHETICAL: NO
; US-09-082-253-5

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x US-09-082-253-5
Align seg 1/1 to: US-09-082-253-5 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTTTCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyLysGluLeuGluPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACCTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTGA 336
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-082-253-5

seq_documentation_block:
; Sequence 5, Application PC/TUS9608235
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08235
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 964796
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; US-09-082-253-5

```



```

; ANTI-SENSE: NO
PCT-US96-08235-5

alignment_scores:
  Quality: 208.00      Length: 73
  Ratio: 3.355        Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x PCT-US96-08235-5 ..
Align seg 1/1 to: PCT-US96-08235-5 from: 1 to: 403

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSerA 35
168 AGTGCTCAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spLaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATCCCACTACAAATGCCATAGATGATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAAACTCTGACCAATGTGAGGTGTTTATGCAATATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CACGAGTCTTTGTGATTGA 336

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-455-896-1

seq_documentation_block:
; Sequence 1, Application US/08455896
; Patent No. 5668267
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,896
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-455-896-1

alignment_scores:
  Quality: 208.00      Length: 73
  Ratio: 3.355        Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:
US-09-367-009-3 x US-08-455-896-1 ..
Align seg 1/1 to: US-08-455-896-1 from: 1 to: 503

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSerA 35
168 AGTGCTCAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 spLaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATCCCACTACAAATGCCATAGATGATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAAACTCTGACCAATGTGAGGTGTTTATGCAATATATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CACGAGTCTTTGTGATTGA 336

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-933-149-1

seq_documentation_block:
; Sequence 1, Application US/08933149
; Patent No. 5922836
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
; TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,149
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: HENDERSON, MELODIE W.
; REGISTRATION NUMBER: 37,848
; REFERENCE/DOCKET NUMBER: 6029-6040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

Copied from PCT009367009 on 27-Feb-2004

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
; US-09-082-253-1
;
; alignment_scores:
; Quality: 208.00 Length: 73
; Ratio: 3.355 Gaps: 0
; Percent Similarity: 84.932 Percent Identity: 53.425
;
; alignment_block:
; US-09-367-009-3 x US-09-082-253-1 ..
;
; Align seg 1/1 to: US-09-082-253-1 from: 1 to: 503
;
; 2 SerGlyCysLysLeuLeuGluAAspMetValGluLysThrIleAsnSerAs 18
; 118 TCTGGCTGCCCTTATTGGAGAGTGTGATTCCTCAAGACATCAATCCACA 167
;
; 18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSerA 35
; 168 AGTGCTCAAGACTGAATACAAAGAACTTCTCAAGAGTTCATAGACGACA 217
;
; 35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
; 218 ATGCCATCAAAATGCCATAGATGAATTGAAGGAATGTTTCTTAACCAA 267
;
; 52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
; 268 ACGGATCAAACTCTGACCAATGTTGAGGTGTTTATGCAATTAATATGA 317
;
; 68 pSerIleTrpCysAsnLeu 74
; 318 CAGCAGTCTTTGTGATTGA 336
;
; seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-08235-1
;
; seq_documentation_block:
; Sequence 1, Application PC/TUS9608235
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEMING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08235
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 964796
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
; PCT-US96-08235-1
;
; alignment_scores:
; Quality: 208.00 Length: 73
; Ratio: 3.355 Gaps: 0
; Percent Similarity: 84.932 Percent Identity: 53.425
;
; alignment_block:
; US-09-367-009-3 x PCT-US96-08235-1 ..
;
; Align seg 1/1 to: PCT-US96-08235-1 from: 1 to: 503
;
; 2 SerGlyCysLysLeuLeuGluAAspMetValGluLysThrIleAsnSerAs 18
; 118 TCTGGCTGCCCTTATTGGAGAGTGTGATTCCTCAAGACATCAATCCACA 167
;
; 18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSerA 35
; 168 AGTGCTCAAGACTGAATACAAAGAACTTCTCAAGAGTTCATAGACGACA 217
;
; 35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
; 218 ATGCCATCAAAATGCCATAGATGAATTGAAGGAATGTTTCTTAACCAA 267
;
; 52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
; 268 ACGGATCAAACTCTGACCAATGTTGAGGTGTTTATGCAATTAATATGA 317
;
; 68 pSerIleTrpCysAsnLeu 74
; 318 CAGCAGTCTTTGTGATTGA 336
;
; seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-947-823-4
;
; seq_documentation_block:
; Sequence 4, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isghouhi
; APPLICANT: Yaghoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Pest Resistance in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California

```

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947.823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-070210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 87..3860
OTHER INFORMATION: /note= "Copy 2 cDNA for M1 nematode
resistance gene of tomato"
1-08-947-823-4

alignment_scores:
Quality: 61.50 Length: 52
Ratio: 1.757 Gaps: 1
Percent Similarity: 67.308 Percent Identity: 28.846

```

align seg 1/1 to: US-08-947-823-4 from: 1 to: 3982
23 GluTyrLysGluLeuGlnGluPheIleAspSerAspAlaAlaG1 39
||| : : : : ||||| : : : : || : : : : || : : : : : : : : : :
321 GAGGTTGAGAATCGTCTCAATCACTTTTGATGATGATGCTCTACTAG 370
39 uAlaMetGlyLysPheLysGlnCysPheLeuAsnGlnSerHisargThrL 56
: : : : : : : : : : : : : : : : ||||| : : : :
371 CCTCACCAAGTAATGATGATGATGATGATGATGATGATGATGCTTCTT 417
56 euLysAsnPheGlyLeuMetMethIsthrValTyrAspSerIletrpCys 72
||||| : : : : : : : : : : ||||| : : : :
418 ATAAATCAGATGCATCATGATGATGATGATGATGATGATGATGATGCTCTCTG 467
73 AsnLeu 74
|||||
468 AATCTG 473

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-947-823-1
seq_documentation_block:
Sequence 1. Application US/08947823
Patent No. 6114605

```

GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Isquuhi

```

seq_documentation_block:
; Sequence 1, Application US/08695866
; Patent No. 5955644
; GENERAL INFORMATION:
; APPLICANT: Hasty, Paul
; APPLICANT: Lim, Dae-Sik
; TITLE OF INVENTION: KU-DEFICIENT CELLS AND NON-HUMAN
; TITLE OF INVENTION: TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,866
; FILING DATE: 08-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 008535-0011-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-695-866-1

alignment_scores:
Quality: 61.00 Length: 52
Ratio: 2.033 Gaps: 1
Percent Similarity: 57.692 Percent Identity: 32.692

alignment_block:
US-09-367-009-3 x US-08-695-866-1 ..
Align seg 1/1 to: US-08-695-866-1 from: 1 to: 2515

16 AsnSerAspIleProGluTyrLysGluLeuGlnGluPheIl 32
||| ::|||::||| |||||:::|||||
793 AACTTGCTATAAGATTGAGCTATATAATCGATTGACAGGAG.... 837

32 eAspSerAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheL 49
|||||::: ::::
838 .....AAATTTAAGAGAGITGGG 856

49 euAsnGlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThr 65
:: |||||::: ::::
857 TAGTTGTGGAGCGAAGACCGCTAAGAGAGAGATATACAAAAGAGACT 906

66 ValTyr 67
|||||
907 GTTAT 912

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:12:13 ; Search time 12.98 seconds
(without alignments)
109.523 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSCCKLLEDMVEKTIINSDIS.....TLKNFLMHTVYDSTWNL 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	390	99.5	95	3	US-08-821-451A-6
2	390	99.5	95	4	US-09-263-810-6
3	208	53.1	93	1	US-08-455-896-2
4	208	53.1	93	2	US-08-933-149-2
5	208	53.1	93	2	US-08-933-149-2
6	208	53.1	93	3	US-09-082-343-2
7	208	53.1	93	3	US-09-082-253-2
8	164	41.8	95	5	PCT-US96-08235-2
9	164	41.8	95	3	US-08-821-451A-27
10	163	41.6	95	4	US-09-263-810-27
11	163	41.6	95	2	US-08-455-896-7
12	163	41.6	95	2	US-08-933-149-7
13	163	41.6	95	2	US-09-082-343-7
14	163	41.6	95	3	US-09-082-253-7
15	61.5	15.7	1257	4	PCT-US96-08235-7
16	59.5	15.2	114	1	US-08-947-823-5
17	59.5	15.2	114	1	US-08-031-399-3
18	59.5	15.2	114	1	US-08-031-399-6
19	59.5	15.2	114	1	US-08-031-399-12
20	59.5	15.2	114	1	US-08-393-305-3
21	59.5	15.2	114	1	US-08-393-305-6
22	59.5	15.2	114	1	US-08-726-817-3
23	59.5	15.2	114	1	US-08-726-817-6
24	59.5	15.2	114	1	US-08-504-042-3
25	59.5	15.2	114	1	US-08-504-042-6
26	59.5	15.2	114	1	US-08-504-042-12
27	59.5	15.2	114	2	US-08-725-969-3
					US-08-725-969-6

28	59.5	15.2	114	2	US-08-794-524-3	Sequence 3, Appl1
29	59.5	15.2	114	2	US-08-794-524-6	Sequence 6, Appl1
30	59.5	15.2	114	4	US-09-189-193-3	Sequence 3, Appl1
31	59.5	15.2	114	4	US-09-189-193-6	Sequence 6, Appl1
32	59.5	15.2	114	5	PCT-US94-03793-3	Sequence 3, Appl1
33	59.5	15.2	114	5	PCT-US94-03793-6	Sequence 6, Appl1
34	59.5	15.2	114	5	PCT-US94-03793-12	Sequence 12, Appl1
35	59.5	15.2	122	1	US-08-300-903A-3	Sequence 3, Appl1
36	59.5	15.2	162	1	US-08-031-399-2	Sequence 2, Appl1
37	59.5	15.2	162	1	US-08-031-399-5	Sequence 5, Appl1
38	59.5	15.2	162	1	US-08-393-305-2	Sequence 2, Appl1
39	59.5	15.2	162	1	US-08-393-305-5	Sequence 5, Appl1
40	59.5	15.2	162	1	US-08-535-733-2	Sequence 2, Appl1
41	59.5	15.2	162	1	US-08-284-393B-9	Sequence 9, Appl1
42	59.5	15.2	162	1	US-08-726-817-2	Sequence 2, Appl1
43	59.5	15.2	162	1	US-08-726-817-5	Sequence 5, Appl1
44	59.5	15.2	162	1	US-08-504-042-2	Sequence 2, Appl1
45	59.5	15.2	162	1	US-08-504-042-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-821-451A-6
; Sequence 6, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-821-451A-6

Query Match 99.5% Score 390; DB 3; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 60
Db 19 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 78
Oy 61 LMMHTVYDYSIWCNL 74
Db 79 LMMHTVYDYSIWCNM 92

RESULT 2
US-09-263-810-6
Sequence 6, Application US/09263810
Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263.810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821.451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-263-810-6

Query Match 99.5%; Score 390; DB 4; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 60
Db 19 DSGCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 78
Oy 61 LMMHTVYDYSIWCNL 74
Db 79 LMMHTVYDYSIWCNM 92

RESULT 3
US-08-455-896-2
Sequence 2, Application US/08455896
Patent No. 568267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455.896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-2

Query Match 53.1%; Score 208; DB 1; Length 93;
Best Local Similarity 53.4%; Pred. No. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
Qy 2 SGCCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 61
Db 20 SGCCKLLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFCQCFLNQSHRTLNKFG 79
Oy 62 LMMHTVYDYSIWCNL 74
Db 80 FMOLIYDSSLCDL 92

RESULT 4
US-08-933-149-2
Sequence 2, Application US/08933149
Patent No. 592836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.2.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELODIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
-08-933-149-2

```

Query Match          53.1%; Score 208; DB 2; Length 93;
Best Local Similarity 53.4%; Pred. No. 1,2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps

```

ULT 5
 09-082-343-2
 sequence 2, Application US/09082343
 Patent No. 5968754
 GENERAL INFORMATION:
 APPLICANT: WATSON, MARK A.
 APPLICANT: FLEMING, TIMOTHY P.
 TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
 TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ROGERS, HOWELL & HAFFERKAMP
 STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
 CITY: ST. LOUIS
 STATE: MISSOURI
 COUNTRY: USA
 ZIP: 63105-1817
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/082,343
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/455,896
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 952726
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (314) 727-5188
 TELEFAX: (314) 727-6092
 INFORMATION FOR SEQ ID NO: 2:

```
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ; HYPOTHETICAL: NO  
US-09-082-343-2
```

Query Match 53.1% Score 208; DB 2; Length 93;
Best Local Similarity 53.4%; Pred. No. 1,2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels

```
QY      2   SGCKLLEDVMEKTTNSDISIPEYKELLQEFDSDAAAEAMGKFCEFLNQSHRT  
| | | | : : : | : | | | | | : | : | : | : |  
Db     20   SGCPLLENYSKTINKPQSKEYKELLQEFDIDNATTNAIDELKECFLENOTDET  
  
QY      62   MMHTVDYSIWCNL 74  
|          ||||| : |  
Db     80   FMOLIYDSSLCDL 92
```

```

RESULT      6
US-09-082-253-2
; Sequence 2, Application US/09082253
; Patent No. 6004756
; GENERAL INFORMATION:
; APPLICANT: WATSON, MARK A.
; APPLICANT: FLEWING, TIMOTHY P.
; TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
; TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082.253
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/455,896
; FILING DATE: 05/31/1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 952726
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; PS-09-082-253-2

```

Query Match 53.1%; Score 208; DB 3; Length 93;
Best Local Similarity 53.4%; Pred. NO. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

us-09-367-009-3.ra1

Tue May 1 11:45:25 2001

QY 2 SGCKLLEDVMTKINSIDISPEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 61
DB 20 SGCPLEENVISKTINPQVSKTEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 79
QY 62 MMHTVYDVSINCL 74
DB 80 FMOLIYDSSICDL 92

RESULT 7
US-09-367-009-3-2
Sequence 2, Application PC/TUS9608235
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAVERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08235
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 964796
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-367-009-3-2

QY 2 SGCKLLEDVMTKINSIDISPEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 61
DB 20 SGCPLEENVISKTINPQVSKTEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 79
QY 62 MMHTVYDVSINCL 74
DB 80 FMOLIYDSSICDL 92

RESULT 8
US-08-821-451A-27
Sequence 2, Application US/08821451A
Patent No. 6066724
GENERAL INFORMATION:
APPLICANT: JIAN NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS

Query Match 53.1%; Score 208; DB 5; Length 93;
Best Local Similarity 53.4%; Pred. No. 1.2e-19;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 2 SGCKLLEDVMTKINSIDISPEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 61
DB 20 SGCPLEENVISKTINPQVSKTEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 79
QY 62 MMHTVYDVSINCL 74
DB 80 FMOLIYDSSICDL 92

Query Match 41.8%; Score 164; DB 3; Length 95;
Best Local Similarity 36.6%; Pred. No. 5.5e-14;
Matches 26; Conservative 19; Mismatches 19; Indels 0; Gaps 0;
QY 2 SGCKLLEDVMTKINSIDISPEYKELLQEFIDSDAAAEAMGKFKCFLNQSHRTLNKFL 61
DB 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQAHFTEKAVKQKOCFLDQTDKTLNVEV 80
QY 62 MMHTVYDVSINCL 72
DB 81 MMEAFNSES 91

RESULT 9
US-09-263-810-27
Sequence 27, Application US/09263810
Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: JIAN NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS

```

; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA: US/09/263.810
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/821,451
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-263-810-27

Query Match 41.8%; Score 164; DB 4; Length 95;
Best Local Similarity 36.6%; Pred. No. 5.5e-14;
Matches 26; Conservative 26; Mismatches 19; Indels 0; Gaps 0;

QY 2 SGCKLLEDVMEKTNISDISIPEYKELQEFIDSDAAAMGKFKQCFLNQSHRTLKNEGL 61
DB 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQAHFTEKAVKQKQCFLDQTDKTLNENGV 80
QY 62 MMHTVYDSIWC 72
DB 81 MMEAFNSESC 91

RESULT 10
US-08-455-896-7
Sequence 7, Application US/08455896
Patent No. 5668267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-7

Query Match 41.6%; Score 163; DB 2; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.4e-14;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 2 SGCKLLEDVMEKTNISDISIPEYKELQEFIDSDAAAMGKFKQCFLNQSHRTLKNEGL 61
DB 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQAHFTEKAVKQKQCFLDQTDKTLNENGV 80
QY 62 MMHTVYDSIWC 72
DB 81 MMEAFNSESC 91

RESULT 11
US-08-933-149-7
Sequence 7, Application US/08933149
Patent No. 5922836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELODIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-933-149-7

Query Match 41.6%; Score 163; DB 2; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.4e-14;
Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 2 SGCKLLEDVMEKTNISDISIPEYKELQEFIDSDAAAMGKFKQCFLNQSHRTLKNEGL 61
DB 21 SGCSILDEVIRGTINSTVTLHDYMKLVKPYQAHFTEKAVKQKQCFLDQTDKTLNENGV 80
QY 62 MMHTVYDSIWC 72
DB 81 MMEAFNSESC 91
```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18602
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.

Search completed: April 24, 2001, 15:13:48
Job time: 95 sec

us-09-367-009-3.rai

Tue May 1 11:45:25 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:11:53 ; Search time 20.75 Seconds
(without alignments)
203.859 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGGKLLDMVEKTIINSIDIS.....TLKNFGLMMHTVYDSIWCNL 74

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

us-09-367-009-3.rag

RESULT 1	
W61649	W61649 standard; peptide; 74 AA.
ID	W61649
XX	XX
AC	W61649;
XX	XX
DT	27-OCT-1998 (first entry)
DE	Non-ocular disease marker 3.
XX	XX
KW	Human; non-ocular disease; tear; cancer; breast; prostate.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO9835229-A1.
XX	XX
PD	13-AUG-1998.
XX	XX
PF	06-FEB-1998; 98WO-AU000071.
XX	XX
PR	07-FEB-1997; 97AU-0005009.
XX	XX
PA	(MACQ-) MACQUARIE RES LTD.
XX	(UNIX) UNISEARCH LTD.
PA	Bolis S, Gooley AA, Herbert B, Molloy M, Morris C;
XX	Walsh B, Willcox M, Williams KL;
PI	WPI; 1998-447373/38.
XX	XX
DR	Screening for non-ocular disease - by analysing tears for marker
XX	proteins, particularly indicative of cancer and genetic disease,
PT	also new proteins and nucleic acid encoding them
XX	XX

ALIGNMENTS	
RESULT 1	
W61649	W61649 standard; peptide; 74 AA.
ID	W61649
XX	XX
AC	W61649;
XX	XX
DT	27-OCT-1998 (first entry)
DE	Non-ocular disease marker 3.
XX	XX
KW	Human; non-ocular disease; tear; cancer; breast; prostate.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO9835229-A1.
XX	XX
PD	13-AUG-1998.
XX	XX
PF	06-FEB-1998; 98WO-AU000071.
XX	XX
PR	07-FEB-1997; 97AU-0005009.
XX	XX
PA	(MACQ-) MACQUARIE RES LTD.
XX	(UNIX) UNISEARCH LTD.
PA	Bolis S, Gooley AA, Herbert B, Molloy M, Morris C;
XX	Walsh B, Willcox M, Williams KL;
PI	WPI; 1998-447373/38.
XX	XX
DR	Screening for non-ocular disease - by analysing tears for marker
XX	proteins, particularly indicative of cancer and genetic disease,
PT	also new proteins and nucleic acid encoding them
XX	XX

SUMMARIES	
Pred. No.	Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
Result No.	Score Query Match Length DB ID Description
1	392 100.0 74 19 W61649 Non-ocular disease
2	390 99.5 95 18 W35804 Human endometrial
3	390 99.5 95 20 Y02590 A human mamoglobin
4	390 99.5 95 21 B03769 Human endometrial
5	390 99.5 95 21 Y92226 Human endometrial
6	390 99.5 95 21 Y92237 Mamoglobin homolo
7	390 99.5 95 21 Y65394 Human 5' EST relat
8	390 99.5 108 20 Y60038 Human endometrium
9	208 53.1 74 21 Y84624 Amino acid sequenc
10	208 53.1 93 18 W10179 Mammary-specific s
11	208 53.1 93 19 W59777 Amino acid sequenc

PS Claim 6; Page 9; 14pp; English.
XX The markers W61647-W61649 are used for screening for, or detecting,
CC non-ocular disease by analysing tears. Biochemicals, specifically
CC proteins, are isolated from tears, particularly by chromatography or
CC electrophoresis, especially two-dimensional polyacrylamide gel
CC electrophoresis (2D-PAGE), then detected, e.g. with labelled specific
CC reagents, in (radio)immunoassay. The method is used to detect cancer,
CC particularly of breast or prostate, or a genetic disease, in humans or
CC animals.

XX Sequence 74 AA;
Query Match 100.0%; Score 392; DB 19; Length 74;
Best Local Similarity 100.0%; Pred. No. 3.8e-41;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVVEKTSINSDISIPYKELLQEFIDSDAAEAMGKFKQCLNOSHRTLNFG 60
|||||
1 dsgcklledmvektinsdisipyeikellqefidsdaaamgkfkqclngshrtlnkf 60
61 LMHTVYDTSIWCNL 74
|||||
61 lmmhtvydsiwcnl 74

RESULT 2
W35804
W35804 standard; Protein; 95 AA.
W35804;

27-MAR-1998 (first entry)
Human endometrial specific steroid-binding factor III.
Endometrial specific steroid-binding factor III; ESF III; human;
Clara cell secretory protein; endometrium;
phospholipase A2 inhibitor; polychlorinated biphenyl; antiaggregant;
inflammation; asthma; rhinitis; cystic fibrosis; airway disease;
neoplasia; atopy; therapy; diagnosis.

Homo sapiens.
Key Location/Qualifiers
Peptide 1..21
Protein 22..95
/label= Sig_peptide
/label= Mat_protein
W09734997-A1.
25-SEP-1997.
21-MAR-1996; 96WO-US03857.
21-MAR-1996; 96WO-US03857.
(HUMA-) HUMAN GENOME SCI INC.
Gentz RL, NI J, Yu G;
WPI; 1997-480206/44.
N-PSDB; T94832.
Human endometrial specific steroid-binding factor I, II and III -
used to treat inflammation, asthma, rhinitis, cystic fibrosis,
airway disease, neoplasia, atopy etc.
Claim 19; Page 65; 92pp; English.
This sequence comprises human endometrial specific steroid binding

CC factor III (ESF III), a protein that inhibits phospholipase A2
CC activity, binds to polychlorinated biphenyl compounds, reduces
CC foreign protein antigenicity, inhibits monocyte and neutrophil
CC chemotaxis and phagocytosis, inhibits platelet aggregation,
CC regulates eicosanoid levels in the human uterus and controls the
CC growth of endometrial cells. The amino acid sequence was deduced
CC from a cDNA clone (see T94832) derived from a human endometrial
CC tumour. ESF I (see W35802) and ESF II (see W35803) are also
CC claimed. Human ESF III has about 36% identity with rat prostag
CC steroid-binding protein. Recombinant ESF I, II and III can be
CC expressed in host cells for use in claimed methods (a) for treating
CC a patient in need of ESF I, II or III (including expression of the
CC polypeptide in vivo) and (b) for identifying compounds which bind
CC to and inhibit activation of the ESF polypeptide. ESF I, II and
CC III may be used to treat inflammation, asthma, rhinitis, cystic
CC fibrosis, airway disease, neoplasia and atopy.

XX Sequence 95 AA;

Query Match 99.5%; Score 390; DB 18; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGCKLLEDVVEKTSINSDISIPYKELLQEFIDSDAAEAMGKFKQCLNOSHRTLNFG 60
Db 19 dsgcklledmvektinsdisipyeikellqefidsdaaamgkfkqclngshrtlnkf 78
QY 61 LMHTVYDTSIWCNL 74
Db 79 lmmhtvydsiwcnm 92

RESULT 3
Y02590
ID Y02590 standard; Protein; 95 AA.
XX Y02590;
AC Y02590;
XX 26-JUL-1999 (first entry)
XX A human mammoglobin homologue (HMH).
XX Human mammoglobin homologue (HMH); antagonist; neoplastic disorder;
KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KW teratocarcinoma; endometriosis.
OS Homo sapiens.
XX W09919487-A1.
XX 22-APR-1999.
XX 14-OCT-1998; 98WO-US21729.
XX 16-OCT-1997; 97US-0951750.
XX (INCY-) INCYTE PHARM INC.
XX Hillman JL, Murry LE, Shah P;
XX WPI; 1999-302531/25.
XX N-PSDB; X36138.
XX New human mammoglobin homologue (HMH), useful for diagnosing, treating
XX or preventing disorders associated with expression of HMH
XX Claim 1; Fig 1A-B; 63pp; English.
XX The present sequence represents a human mammoglobin homologue (HMH).
CC Antagonists of the HMH polypeptide can be used to treat neoplastic
CC disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
CC myeloma, sarcoma and teratocarcinoma. A vector expressing the

complement of the polynucleotide encoding HMH may be administered to a subject to treat or prevent neoplastic disorders or endometriosis. Antibodies which bind HMH may also be used in the diagnosis of conditions or diseases characterized by expression of HMH, or in assays to monitor patients being treated with HMH, agonists, antagonists or inhibitors. Polynucleotides encoding HMH may also be used diagnostically to detect and quantitate gene expression in biopsied tissues. With respect to cancer a relatively high amount of transcript may indicate a predisposition for the development of disease. The nucleic acid sequences which encode HMH may also be used to generate hybridization probes useful for mapping the naturally occurring genomic sequence. HMH, and its fragments/variants can be used for screening libraries of compounds in drug screening techniques.

Sequence 95 AA;

Query Match 99.5%; Score 390; DB 20; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 DSGCKLLEDVMEKTIINSISPEYKELQEFIDSDAAAMGKFKQCFNLNOSHRCLKNFG 60
|||||
19 dsgcklledvmeaktinsdisipeykellqefidsdaaamgkfkqcfnlngshrtlkngf 78

61 LMMHTVYDSIWCNL 74
|||||
79 lmmhtvydsiwcnm 92

RESULT 4

B03769
B03769 standard; Protein; 95 AA.

B03769;

06-OCT-2000 (first entry)

Human endometrial specific steroid-binding factor III protein sequence.

Endometrial specific steroid-binding factor; human; hESF; inflammation; asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy; eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.

Homo sapiens.

US6066724-A.

23-MAY-2000.

21-MAR-1997; 97US-0821451.

21-MAR-1996; 96US-0014724.

(HUMA-) HUMAN GENOME SCI INC.

Yu G, Gentz R, Ni J;

WPI; 2000-375600/32.

N-PSDB; A59730.

Novel gene encoding human endometrial specific steroid-binding factor I, II and III which is useful for treating asthma, rhinitis, cystic fibrosis, airway disease and neoplasia

Claim 1; Fig 3; 36pp; English.

This invention relates to nucleic acid molecules encoding portions of the human endometrial specific steroid-binding factors I, II, and III. Also included in the invention are hESF I, II, and III polypeptide sequences. The nucleotide sequence exhibit antiasthmatic, antiinflammatory, antiallergic, and cytostatic properties. The polynucleotides are used in gene therapy to express hESF I, II and III polypeptides in vivo to treat

and/or prevent inflammation, asthma, rhinitis, cystic fibrosis, air way disease, neoplasia and atopy. The polynucleotides are also used to inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis and phagocytosis, inhibit platelet aggregation, regulate eicosanoid levels in the human uterus and control the growth of endometrial cells. The polynucleotides are also useful for detecting complementary polynucleotides as a diagnostic reagent. The hESF I, II and III polynucleotides are used to detect complementary polynucleotides such as a diagnostic reagent. Detection of a mutated form of hESF I, II and III associated with a dysfunction will provide a diagnostic tool that can define diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of hESF I, II and III e.g. a susceptibility to inherited asthma and endometrial cancer. They are also useful for chromosome identification. The present sequence represents a hESF III protein sequence identified in the invention.

Sequence 95 AA;

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DSGCKLLEDVMEKTIINSISPEYKELQEFIDSDAAAMGKFKQCFNLNOSHRCLKNFG 60
|||||
Db 19 dsgcklledvmeaktinsdisipeykellqefidsdaaamgkfkqcfnlngshrtlkngf 78

Oy 61 LMMHTVYDSIWCNL 74

|||||

Db 79 lmmhtvydsiwcnm 92

RESULT 5

Y92226
ID Y92226 standard; Protein; 95 AA.

AC Y92226;

10-AUG-2000 (first entry)

Human endometrial specific steroid binding factor III.

Endometrial specific steroid binding factor; ESBPIII; diagnosis; gynaecological cancer; uterine; breast; endometrial; ovarian; antibody; gene therapy; cytostatic.

Homo sapiens.

WO200020044-A1.

13-APR-2000.

30-SEP-1999; 99WO-US222753.

02-OCT-1998; 98US-0102743.

(DIAD-) DIADEXUS LLC.

Macina RA;

WPI; 2000-303649/26.

N-PSDB; A09104.

Diagnosing, staging and monitoring gynecological cancer comprising using an elevated level of ESBPIII in a patient as an indicator of cancer

Claim 6; Page 28-29; 32pp; English.

The levels of human endometrial specific steroid binding factor (ESBPIII) can be measured and compared to control levels and used to

CC diagnose the presence of a gynaecological (uterine, breast, endometrial,
CC or ovarian) cancer in a patient. ESBPIII levels can also be used to
CC diagnose metastasis, to stage or monitor gynaecological cancer.
CC Antibodies specific for ESBPIII can be used to treat gynecological
CC cancers.
XX
SQ Sequence 95 AA;

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 DSGCKLLEDVVEKTSINSDISPEYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNFG 60
19 dsgcklledmvektinsdispeykellqefidsdaaaeamgkfkqcflnqshrtlnfg 78
61 LMHTVYDSIWCNL 74
79 lmmhtvydsiwcnm 92

RESULT 6
Y92237
Y92237 standard; Protein; 95 AA.
Y92237;
10-AUG-2000 (first entry)
Mammoglobin homologue from clone Mamm-X.
Clone Mamm-X; mammaglobin; breast cancer; cytostatic; anti-HIV;
immunosuppressive; antiallergic; antineoplastic; antiinflammatory;
antiarthritic; antiarteriosclerotic; vasotropic; neuroprotective;
nootropic; dermatological; tranquilizer; vulnerary.

Homo sapiens.
WO200020447-A2.
13-APR-2000.
06-OCT-1999; 99WO-US23294.
06-OCT-1999; 98US-0103195.
05-OCT-1999; 99US-0103195.
(CURA-) CURAGEN CORP.
Shinkets RA;
WPI; 2000-303741/26.
N-PSDB; A09118.
Nucleic acids encoding polypeptides with syncollin-like, claudin-like or
cytokine-like activity, useful for treating diseases including cancer,
Alzheimer's and atherosclerosis

Claim 23; Fig 6; 118pp; English.
Clone Mamm-X encodes a polypeptide that is 100 percent identical to human
Mammaglobin B precursor, a potential marker of breast cancer nodal
metastasis. The sequences are useful for treatment of diseases such as
cancer, immune disorders, autoimmune disease, acquired immune deficiency
syndrome (AIDS), transplant rejection, allergy, infection by a
pathological agent or organism, inflammatory disorders, arthritis, a
haematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a
neurological disease, Alzheimer's disease, trauma, spinal cord injury
and skeletal disorders.

Sequence 95 AA;
XX

Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 DSGCKLLEDVVEKTSINSDISPEYKELLOEFIDSDAAAEAMGKQCFLNQSHRTLNFG 60
19 dsgcklledmvektinsdispeykellqefidsdaaaeamgkfkqcflnqshrtlnfg 78
61 LMHTVYDSIWCNL 74
79 lmmhtvydsiwcnm 92

RESULT 7
Y65394
Y65394 standard; Protein; 95 AA.
XX
AC Y65394;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST related polypeptide SEQ ID NO.1555.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.
XX
OS Homo sapiens.
XX
PN WO9953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
XX
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.

XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
XX
DR N-PSDB; 243008.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
XX
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 3; Page 818; 837pp; English.
XX
XX
CC 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y65438
CC represent the EST-related proteins corresponding to 242265 to 243052.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.
XX

SQ Sequence 95 AA;
Query Match 99.5%; Score 390; DB 21; Length 95;
Best Local Similarity 98.6%; Pred. No. 9.3e-41;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHTLKNFG 60
DB 19 dsgcklledvktinsdisipeykellqefidsdaaaamgkfkqcflngshrtlkngf 78
QY 61 LMMHTVYDSIWCNL 74
DB 79 lmmhtvydsiwcnm 92
RESULT 8
ID Y60038 standard; Protein; 108 AA.
XX Y60038;
XX 31-JAN-2000 (first entry)
XX Human endometrium tumour EST encoded protein 98.
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
XX treatment; uterine; gene therapy; expressed sequence tag.
XX Homo sapiens.
XX DE19817948-A1.
XX 21-OCT-1999.
XX 17-APR-1998; 98DE-1017948.
XX 17-APR-1998; 98DE-1017948.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
XX WPI; 1999-591957/51.
XX N-PSDB; 242013.
XX New nucleic acid sequences expressed in uterine cancer tissues, and
XX derived polypeptides, for treatment of uterine and endometrial cancer
XX and identification of therapeutic agents -
XX Claim 23; Page 314; 444pp; German.
XX This invention describes novel human nucleic acid (cDNA) sequences (A),
XX that are highly expressed in uterine tumour tissue and which have
XX anticancer and cytostatic activity. (A) are used (i) for recombinant
XX expression of polypeptides (B) and (ii) to isolate complete genes. (B)
XX are used (i) to identify agents suitable for treatment of uterine or
XX endometrial cancer; (ii) directly for treating these forms of cancer
XX (including expression from gene therapy vectors) and (iii) for generation
XX of specific antibodies. (A) are identified by assembling ESTs (expressed
XX sequence tags) from a particular tissue type before comparison of
XX expression patterns. This allows a significantly longer fragment of the
XX gene to be revealed, so should reduce the number of failures associated
XX with the fact that ESTs from different libraries may represent different
XX parts of the same unknown gene, distorting the estimated frequency of
XX occurrence in a particular tissue. X59941-V60328 represent protein
XX fragments encoded by the human endometrium tumour cDNA library derived
XX EST fragments represented in Z41981-Z42121.
XX Sequence 108 AA;

Query Match 99.5%; Score 390; DB 20; Length 108;

Best Local Similarity 98.6%; Pred. No. 1.1e-40;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGCKLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHTLKNFG 60
DB 32 dsgcklledvktinsdisipeykellqefidsdaaaamgkfkqcflngshrtlkngf 91
QY 61 LMMHTVYDSIWCNL 74
DB 92 lmmhtvydsiwcnm 105
RESULT 9
ID Y84624 standard; Protein; 74 AA.
XX Y84624;
XX 25-JUL-2000 (first entry)
XX Amino acid sequence of mammary-specific protein mammaglobin fragment.
XX Human; mammaglobin; mammary gland; breast cancer; endometrial cancer.
XX Homo sapiens.
XX WO200018783-A1.
XX 06-APR-2000.
XX 29-SEP-1999; 99WO-US22616.
XX 29-SEP-1998; 98US-0162622.
XX (UNIW) UNIV WASHINGTON.
XX Watson MA, Fleming TP;
XX WPI; 2000-293105/25.
XX Methods for detecting breast cancer, comprising detecting elevated
XX concentrations of a mammaglobin polypeptide, using an antibody, or
XX detecting elevated concentrations of the mRNA encoding the polypeptide,
XX using oligonucleotides -
XX Claim 14; Page 67-68; 71pp; English.
XX The present sequence represents a fragment of human mammary-specific
XX secreted protein, mammaglobin. Mammaglobin expression is restricted
XX to the mammary gland. Dysregulation of the mammaglobin gene occurs early
XX and frequently in breast cancer. The specification describes a method
XX for detecting the presence of breast cancer in a patient, comprising
XX detecting an elevated concentration of mRNA encoding a mammaglobin
XX polypeptide. The methods are useful for detecting the presence of
XX breast and endometrial cancer.
XX Sequence 74 AA;

Query Match 53.1%; Score 208; DB 21; Length 74;
Best Local Similarity 53.4%; Pred. No. 1.9e-18;
Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 2 SGCKLEDVKTINSISIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHTLKNFG 61
DB 1 sgcklledvktinsdisipeykellqefidsdaaaamgkfkqcflngshrtlkngf 60
QY 62 MMHTVYDSIWCNL 74
DB 61 fmqliydsiwdl 73
RESULT 10


```

XX 29-SEP-1999; 99WO-US22616.
XX PF
XX 29-SEP-1998; 98US-0162622.
XX PR
XX 29-SEP-1998; 98US-0162622.
XX PA
XX (UNIW) UNIV WASHINGTON.
XX PA
XX Watson MA, Fleming TP;
XX PI
XX 2000-293105/25.
XX WI
XX N-PSDB; A12632.
XX DR
XX
XX Methods for detecting breast cancer, comprising detecting elevated
XX concentrations of a mammaglobin polypeptide, using an antibody, or
XX detecting elevated concentrations of the mRNA encoding the polypeptide,
XX using oligonucleotides -
XX
XX Example 1; Fig 2; 71pp; English.
XX
XX The present sequence represents the human mammary-specific secreted
XX protein mammaglobin. Mammaglobin expression is restricted to the
XX mammary gland. Dysregulation of the mammaglobin gene occurs early
XX and frequently in breast cancer. The specification describes a method
XX for detecting the presence of breast cancer in a patient, comprising
XX detecting an elevated concentration of mRNA encoding a mammaglobin
XX polypeptide. The methods are useful for detecting the presence of
XX breast and endometrial cancer.
XX
XX Sequence 93 AA;
XX
XX Query Match 53.1%; Score 208; DB 21; Length 93;
XX Best Local Similarity 53.4%; Pred. No. 2.5e-18;
XX Mismatches 39; Conservative 12; Mismatches 22; Indels 0; Gaps
XX
XX 2 SGCKLEEDVVEKTIINSDISIPEYKELLOEFTDSDAAFAMGKFKQCFLNQSHRTLNKFLG 61
XX ||| |||||: ||||: ||||| |||||: ||: |||||: |||: ||:
XX 20 sgcpllenvisktipqvskvkeykellqefidnnaidclqcdetclnsnvev 79
XX
XX 62 MMHTVYDSIWCNL 74
XX | : ||| ||: |
XX 80 fmqliydsslcld 92
XX
XX RESULT 15
XX B51127
XX B51127 standard; Protein; 93 AA.
XX
XX B51127;
XX
XX 20-MAR-2001 (first entry)
XX
XX Human mammaglobin amino acid sequence SEQ ID NO:27.
XX
XX Human; mammaglobin; breast cancer; detection; diagnosis; antibody;
XX vaccine; cytostatic; antimammaglobin.
XX
XX Homo sapiens.
XX
XX WO200073338-A1.
XX
XX 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14845.
XX
XX 28-MAY-1999; 99US-0136528.
XX
XX 01-JUN-1999; 99US-0137048.
XX
XX (CORI-) CORIXA CORP.
XX
XX (HEND-) HENDRICKSON R C.
XX
XX (HOUG-) HOUGHTON R L.
XX
XX (REED/) REED S G.
XX

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:12:57 ; Search time 13.62 Seconds
(without alignments)
373.385 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSCCKLLEDVMTKINSDIS.....TLKNFGLMHMTVYDSIWCNL 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-67.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	41.6	95	1 BORT3	prostatic steroid-
2	76.5	19.5	280	2 T38449	hypothetical prote
3	70	17.9	303	2 B71638	cell division prot
4	66.5	17.0	234	2 F64104	hypothetical prote
5	65.5	16.7	131	2 F70880	El protein - human
6	65.5	16.7	603	1 W1WL8	hypothetical prote
7	64.5	16.5	153	2 E82928	dynam-lik-like prote
8	64.5	16.5	610	2 S59558	hypothetical prote
9	63.5	16.2	269	2 G75148	hypothetical prote
10	63	16.1	131	2 D81908	hypothetical prote
11	62.5	15.9	496	2 T46356	hypothetical prote
12	62.5	15.9	2485	1 H71621	serine/threonine-s
13	62	15.8	131	2 D81108	conserved hypothet
14	62	15.8	764	2 H71607	hypothetical prote
15	62	15.8	2013	2 C71610	probable membrane
16	61.5	15.7	162	2 I49124	interleukin-15 - m
17	61.5	15.7	734	2 E72271	5-methyltetrahydro
18	61.5	15.7	1257	2 T06269	root-knot nematode
19	61	15.6	490	2 T31646	hypothetical prote
20	61	15.6	605	2 S36469	El protein - human
21	61	15.6	663	2 T37772	telomere length re
22	61	15.6	732	2 S26303	Ku autoantigen 80K
23	60.5	15.4	269	2 T71188	hypothetical prote
24	60.5	15.4	430	2 T04668	phosphoserine tran
25	60.5	15.4	606	2 S60090	rad17 protein - fi
26	60.5	15.4	611	2 S61147	TCM10 protein - ye
27	60.5	15.4	1213	2 T41378	probable helicase
28	60.5	15.4	1601	2 T18800	hypothetical prote
29	60	15.3	268	1 GWKECC	indole-3-glycerol-

30 60 15.3 534 2 A37483 F protein - Muraya
31 60 15.3 606 1 W1WL85 El protein - human
32 60 15.3 790 1 WMFM9 90K protein - alfa
33 60 15.3 1071 2 S38164 ATP-binding protei
34 60 15.3 3795 2 T00831 hypothetrical prote
35 59.5 15.2 162 1 A53484 interleukin-15 pre
36 59.5 15.2 312 2 E70376 exopoliphosphatase
37 59.5 15.2 785 2 D71653 cell surface anig
38 59.5 15.2 925 2 T37831 hypothetrical prote
39 59.5 15.2 1075 2 S54067 probable membrane
40 59.5 15.2 1177 2 I64233 hypothetrical prote
41 59.5 15.2 1198 2 T20262 hypothetrical prote
42 59.5 15.2 1255 2 T06267 nenatodes resistan
43 59 15.1 173 2 T01889 hypothetrical prote
44 59 15.1 604 2 S36540 El protein - human
45 59 15.1 606 1 W1WL5 El protein - human

ALIGNMENTS

RESULT 1

BORT3
prostatic steroid-binding protein chain C3 precursor - rat
N:Alternate names: prostatin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Dec-1982 #sequence_revision 14-Nov-1993 #text_change 28-May-1999
C:Accession: A92395; A92433; A91108; A42392; A03250
R:Parker, M.G.; White, R.; Hurst, H.; Needham, M.; Tilly, R.
J Biol. Chem. 258, 12-15, 1983
A:Title: Prostatic steroid-binding protein. Isolation and characterization of C3 gene
A:Reference number: A92395; MUID:83082848
A:Accession: A92395
A:Molecule type: mRNA
A:Residues: 1-95 <PAR>
A:Cross-references: GB:V01263; GB:J00777; NID:q56993; PIDN:CAA24577.1; PID:q56994
R:Viskochil, D.H.; Perry, S.T.; Lee, O.A.; Stafford, D.W.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 258, 8861-8866, 1983
A:Title: Isolation of two genomic sequences encoding the M-r = 14,000 subunit of rat
A:Reference number: A92433; MUID:83238526
A:Accession: A92433
A:Molecule type: mRNA
A:Residues: 1-78, S', 80-95 <VIS>
R:Peeters, B.; Rombauts, W.; Mous, J.; Heyns, W.
Eur. J. Biochem. 115, 115-121, 1981
A:Title: Structural studies on rat prostatic binding protein. The primary structure o
A:Reference number: A91108; MUID:81188769
A:Accession: A91108
A:Molecule type: protein
A:Residues: 19-95 <PEE>
R:Tan, J.A.; Marschke, K.B.; Ho, K.C.; Perry, S.T.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 267, 4456-4466, 1992
A:Title: Response elements of the androgen-regulated C3 gene.
A:Reference number: A42392; MUID:92165796
A:Accession: A42392
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-16, T', 18-95 <TAN>
A:Note: sequence inconsistent with the nucleotide translation
C:Comment: C3 is encoded by two unique genes that differ from each other only in thei
ng C2 and C3 chains. The chains of each dimer are linked by disulfide bonds. This pro
C:Comment: Steroid-binding protein, the principal androgen-dependent secretory protei
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-95/Product: prostatic steroid-binding protein C3 chain #status experimental <MAT
F:35/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 41.6%; Score 163; DB 1; Length 95;
Best Local Similarity 36.6%; Pred. No. 7.5e-11;

Matches 26; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 2 SCKLLEDVVEKTSINSDISIPYKELLOEFIDSDAAAEAMGKFKOCFLNOSHRTLNKFLG 61

Db 21 SCSILDEVIRGTINSTVTLHDYMKLVKPYQVHFTEKAVKQFKOCFLDQTKLTENGV 80

QY 62 MMHTVYDSIWC 72

Db 81 MWEAIFNSES 91

RESULT 2

F64104

hypothetical protein SPAC27E2.02 - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38449

C:Authors: L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

C:Submitted to the EMBL Data Library, September 1997

C:Reference number: 221749

C:Accession: T38449

C>Status: preliminary; translated from GB/EMBL/DBJ

C:Molecule type: DNA

C:Residues: 1-280 <MUR>

C:Cross-references: EMBL:Z98978; PIDN:CAB11676.1; GSPDB:GN000066; SPDB:SPAC27E2.02

C:Experimental source: strain 972h; cosmid c27E2

C:Genetics:

C:Gene: SPDB:SPAC27E2.02

C:Map position: 1

C:Introns: 185/3; 225/3

Query Match

Best Local Similarity 19.5%; Score 76.5; DB 2; Length 280;

Matches 23; Conservative 10; Mismatches 28; Indels 15; Gaps 3;

QY 6 LLEDVVEKTSINSDISIPYKELLOEFIDSD---AAAEAMGKFKO-----CFLNOSHRT 55

Db 68 LAEDVLLSVATGVCIFSYMDLKLVDIAEQAAERESKLOESDKETPYMLNKSHYV 127

QY 56 LKNFLMMHTVYDSIW 71

Db 128 AKT-----PEIQDEPW 138

RESULT 3

F71638

cell division protein ftsy (ftsY) RP775 - *Rickettsia prowazekii*

C:Species: *Rickettsia prowazekii*

C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000

C:Accession: B71638

C:Authors: S.G.E.; Zomorodipour, A.; Andersson, J.O.; Scharitz-Ponten, T.; Alsmark,

C:Nature 396, 133-140, 1998

C>Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.

C:Reference number: A71630; MUID:99039499

C:Accession: B71638

C>Status: preliminary; nucleic acid sequence not shown; translation not shown

C:Molecule type: DNA

C:Residues: 1-303 <AND>

C:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAAL5202.1; PID:g386130

C:Experimental source: strain Madrid E

C:Genetics:

C:Gene: ftsY; RP775

C:Superfamily: docking protein

Query Match

Best Local Similarity 17.9%; Score 70; DB 2; Length 303;

Matches 19; Conservative 10; Mismatches 17; Indels 8; Gaps 1;

QY 7 LEDVVEKTSINSDISIPYKELLOEF-----IDSDAAAEAMGKFKOCFLNQS 52

Db 37 LNELELLISSDISISVYTHIEEFKNVKKFDITSDTVKIAKLEIQQLSKS 90

RESULT 4

F64104

hypothetical protein HI0952 - *Haemophilus influenzae* (strain Rd KW20)

C:Species: *Haemophilus influenzae*

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999

C:Accession: F64104

C:Authors: R.Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

C:; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman

C:; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

C:Science 269, 496-512, 1995

C:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

C:A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

C:A:Reference number: A64000; MUID:95350630

C:A:Accession: F64104

C:A:Molecule type: DNA

C:A:Status: nucleic acid sequence not shown; translation not shown

C:A:Residues: 1-234 <TIGR>

C:A:Cross-references: GB:U32776; GB:L42023; NID:g1573969; PIDN:AAC22613.1; PID:g1573977

C:C:Genetics:

C:A:Start codon: GTG

C:A:Superfamily: DNA repair protein radc

Query Match

Best Local Similarity 17.0%; Score 66.5; DB 2; Length 234;

Matches 21; Conservative 15; Mismatches 26; Indels 17; Gaps 3;

QY 6 LLEDVVEKTSINSDISIPYKELLOEFIDSDAAAEAMGKFKOCFLNOSHRTLNKFLMHT 65

Db 103 LKQDMLSTPIIND---PETVKL---FLTLEQHEEREVFMVFLDNQHLIKKRLFLGT 156

QY 66 VY-----DSIWCN 73

Db 157 IVVSAYVPREIKEALYCN 175

RESULT 5

F70880

hypothetical protein Rv2759c - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70880

C:Authors: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

C:; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

C:; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

C:Nature 393, 537-544, 1998

C:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

C:A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno

C:A:Reference number: A70500; MUID:98295987

C:A:Accession: F70880

C:A:Status: preliminary; nucleic acid sequence not shown; translation not shown

C:A:Molecule type: DNA

C:A:Residues: 1-131 <COL>

C:A:Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAAL5555.1; PID:ell7

C:A:Experimental source: strain H37RV

C:Genetics:

C:Gene: Rv2759c

Query Match

Best Local Similarity 16.7%; Score 65.5; DB 2; Length 131;

Matches 19; Conservative 13; Mismatches 28; Indels 25; Gaps 1;

QY 1 DSGCKLLEDVVEKTSINSDISIPYKELLO-----EFTDSO 35

Db 15 ESGAOVLKEALERSPNRMSAPNTVELCAIMQRRDRPEISRLVDRLLDDYGVIOVEAYDAD 74

QY 36 AAAEAMGKFKOCFLNOSHRTLNKFG 60

Db 75 QARYAAQAYRDYGRSGHPARLNLG 99

```

RESULT 6
W1WL8
EI protein - human papillomavirus type 8
C:Species: human papillomavirus type 8
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 07-Nov-1997
C:Accession: A03656
R:Fuchs, P.G.; Iftner, T.; Weninger, J.; Pfister, H.
J. Virol. 58, 626-634, 1986
A:Title: Epidermodysplasia verruciformis-associated human papillomavirus 8: genomic sequence
A:Reference number: A93019; MUID:86200410
A:Accession: A03656
A:Molecule type: DNA
A:Residues: 1-603 <FUC>
A:Cross-references: GB:M12737; NID:g333074
A>Note: this ORF is not annotated in GenBank entry PPH8CG
C:Superfamily: papillomavirus EI protein
C:Keywords: early protein

Query Match 16.7%; Score 65.5; DB 1; Length 603;
Best Local Similarity 22.2%; Pred. No. 23;
Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;

Query 1 DSGCKILEDMVEKTNDSISIP-----EYKELQEFIDSDAAAEAMGK 43
Db 112 DSGVELTLNNEADYSHEVEIPAIDSRPEDEGSGALDIDYTALLR---SSNTKATLMK 168
Query 44 FKQCFLN-----QSHRTLKNFGLM---MHTVYDS 69
Db 169 FKFAFGDGFNELTROFKSVKTCNYYWVAAYAVHDVYES 207

RESULT 7
E82928
Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82928
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
Submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: E82928
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-153 <GLA>
A:Cross-references: GB:AE002115; GB:AF222894; NID:g6899102; PIDN:AAF30562.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
A:Genetics:
A:Gene: UU156
A:Genetic code: SGC3

Query Match 16.5%; Score 64.5; DB 2; Length 153;
Best Local Similarity 30.2%; Pred. No. 6.3;
Matches 19; Conservative 13; Mismatches 22; Indels 9; Gaps 3;

Query 4 CKLLEDVVEKTNDS--ISIP EYKELQEFIDSDAAAEAMGKFKQCFLNQSHRT-----L 56
Db 49 CKIVPRMISKHLNNDHLINVDYVPLTKFEIKKSVKEV--KHYLRFISYENKTKRQMPDL 106
Query 57 KNF 59
Db 107 INF 109

RESULT 8
E59558
Glycyl-tRNA synthetase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000

```

A;Residues: 1-131 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84635.1; PID:g738005
A;Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA1393; NMA1394

Query Match 16.1%; Score 63; DB 2; Length 131;
Best Local Similarity 27.5%; Pred. No. 7.7;
Matches 22; Conservative 14; Mismatches 30; Indels 14; Gaps 4;

2 SGCKLLEDMVEKTINSDISIPEYKELL-----QEFIDSDAAE--AMGKFQCF-LNQ 51

```

1 1 1 : 1 : : 1 1 : 1 1 : 1 1 :
51 50CRTLSDLTPENLSDIKPVDYPEILIGTGAACEFIHPKIMAFDSRIGISVECMNTDS 110

```

```

52 SHRTLKNEGLMMHTVYDSIW 71
   : ||| : :|
11 AFRTL---VELHSEGRRAW 126

```

RESULT 11
 6356
 Hypothetical protein DKF2p343G2016.1 - human (fragment)
 Species: Homo sapiens (man)
 Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 Accession: T46356
 Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 Submitted to the Protein Sequence Database, January 2000
 Reference number: 223037

Query Match 15.9%; Score 62.5; DB 2; Length 496;
Best Local Similarity 28.2%; Pred. No. 39;
Matches 20; Conservative 14; Mismatches 20; Indels 17; Gaps 3;

4 CKLLEDNVEKTI NSDISIPEYK-----ELL-----QEFIDSAAAEAMGKFQC 47
|::: |: : : | | | | : | : | | |
CRLLORKMIDISLOGELTTPVOKICKYPLOLAELKYTHPOHRDEKDV EALHMKNVAQ- 269

```

48 FLNQSHRTLKN 58
      :! : |! :!
270 LINERKRRLEN 280

```

RESULT 12
H71621
serine/threonine-specific protein kinase (EC 2.7.1.1-) PFB0150C - malaria parasite (Plasm
C:Species: Plasmodium falciparum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: H71621
R:Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, Q.; Smith, H.Q.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: H71621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2485 <GAR>
A:Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AACT1820.1; PID:g384510
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0150C
C:Superfamily: malaria parasite
C:Keywords: phosphotransferase

```
Query Match      15.8%; Score 62; DB 2; Length 764;
Best Local Similarity 20.8%; Pred. No. 71;
Matches 15; Conservative 19; Mismatches 36; Indels 2; Gaps 1;
```

QY 1 DSCKKLEDMVKETINSIDISPEYKELLOEFIFSDAAAE--AMGKFQCFLNQSHRFLKN 58
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Ddb 149 DDDEICEENLDKHVNDPLNVEKKMLERFNCDECQEITTLNNMKECIVMCMYNFNFS 208

QY 59 FGLMHVTVDYSI 70
 ::::|::|
Db 209 CRIISLLFYNRI 220

RESULT 15
C71610
Probable membrane associated protein PFB0615c - malaria parasite (Plasmodium falciparum)
C-Species: Plasmodium falciparum
C-Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C-Accession: C71610
A-R.Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
A-Porter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
A-Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A-Reference number: A71600; MUID:99021743
A-Accession: C71610
A-Status: preliminary; nucleic acid sequence not shown; translation not shown
A-Molecule type: DNA
A-Residues: 1-2013 <GAR>
A-Cross-references: GB:AF001406; NID:g3845230; PIDN:AAC71912.1; PID:g3845230
A-Experimental source: clone 3D7
A-Gene: PFB0615c

```

Query Match      . 15.8%; Score 62; DB 2; Length 1013;
Best Local Similarity 28.9%; Pred. No. 2.1e+02;
Matches 22; Conservative 4; Mismatches 32; Indels 18; Gaps 1;

Dy 8 EDWVKTINSDISISPEYKELQE-----FIDSDAAAEAMGKQCF 49
   || : || | | | | | | | | | | | | | | | | | |
Db 1158 EDKLEKINEDFVITEEGEKSKIKNTQHDNNNNNDVFCNLSYELLNKEKSF 1217
      | | | | | | | | | | | | | | | | | | | | | |
Dy 50 NOSHETLKNFGLMHT 65
   | | | | | | | | | | | | | | | | | | | | | |
Db 1218 NIKHGKLYINERMHT 1233
      | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: April 24, 2001, 15:14:09
Job time: 72 sec

Job time: 72 sec

ID	Q9QXF3	PRELIMINARY;	PRT;	94 AA.
NC	Q9QYF3.			

April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottler P.,
 Burtis K.C., Bugam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ilegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.-H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*;"
 Science 287:2185-2195(2000).
 [2]
 SEQUENCE FROM N.A.
 Hendrick A.G., Harrington L.S.;
 "Identification of novel ARF-GTPase activating protein (GAP)-like
 proteins in the *Drosophila* genome.";
 Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 EMBL: AF003742; AAF56100.1; -;
 EMBL: AF254071; AAF64529.1; -;
 HSSP: P25098; 1BAK.
 FLYBASE: Fgn0039056; CG6742.
 INTERPRO: IPR001184;
 INTERPRO: IPR001547; -;
 INTERPRO: IPR001849; -;
 INTERPRO: IPR002110; -;
 PFAM: PF00023; ank; 2.
 PFAM: PF00169; ph; 1.
 PFAM: PF01412; ArfGap; 1.
 PRINTS: PRO0405; REVINTRACTNG.
 PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 PROSITE: PS50003; PH_DOMAIN; 1.
 SEQUENCE 828 AA; 93573 MW; 1CED31042512377C CRC64;

QY	1	DSGCKLLEDMVEKTIN-SDISIPYKE-----LLQEPIDSDAAAEAMGFKF 45
Db	25	ESDIEHLEORLEKIICLCNVAVDVGKYYVQSAFAMSLWDIQQHFLDNKNAHNAIGKLI 84
QY	46	OCF-----LNQSHRT-LKNFGLMM 63
Db	85	HCFQEMNKFFHTILLDQASRTVLKNLSVVF 113
RESULT	6	
Q9V7N3		
ID	Q9V7N3	PRELIMINARY; PRT; 1187 AA.
AC	Q9V7N3	
DT	01-MAY-2000	(TREM)rel. 13, Created)

01-WAY-2000 (TRENBLrel. 13, Last sequence update)
01-OCT-2000 (TRENBLrel. 15, Last annotation update)
CG8060 PROTEIN.
CG8060.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=BERKELEY.
MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yeung M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazer G.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.E., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabriellian A.C., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu X., Smith H.O.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*;"
Science 287:2185-2195(2000).
DR EMBL; AF003807; AAF36015.1; -;
DR HSSP; P42771; 1B17.
DR FLYBASE; FBgn0034113; CG8060.
DR INTERPRO; IPR000210; -;
DR INTERPRO; IPR000408; -;
DR INTERPRO; IPR002110; -;
DR PFAM; PF00023; ank; 2.
DR PFAM; PF00651; BTB; 1.
SQ SEQUENCE 1187 AA; 135391 MW; 7E27482BDF2C2A78 CRC64

```

Query Match      17.9%; Score 70; DB 5; Length 1187;
Best Local Similarity 27.4%; Pred. No. 28;
Matches 23; Conservative 9; Mismatches 20; Indels 32; Gaps 3;

QY 20 SIP-----EYKELLQEFIDSDAAAE-----AMGKF-----KQC 47
|:| :|:| | | | : | | | |
537 SLPRBHSFKLLHETSDCAVDHWFVHDGKFAAHKFIYSRAPGLURDLIRCYLDKDI 596

QY 48 FLNQSHRTLKNFGIMMHTVYDSIW 71
|:| | | | :|:| :|:| :|:|
597 YLNFDDLHTGKMFELILNHHIYSSYW 620

```


Copied from PCT009367009 on 27-04-2004

OA	NCBL_taxid=7227;
RN	[1]
RP	SEQUENCE FROM N.A.

Query Match 16.2%; Score 63.5; DB 5; Length 2162;
Best Local Similarity 38.9%; Pred. No. 2.8e+02;
Matches 21; Conservative 8; Mismatches 14; Indels 11; Gaps 3;

QY 5 KLEDVTEKTINSDISITPEYKELLQEFDSDAAAPAMGKFQCFLNQSHRTLKN 58
: | | : | | | : :: | | | | : | | : | | : | | :

D6 1579 FLIADFEKTK-GSKVA- ---KELL-----DNWBAQGKFVKVFPEYEQAKLD 1621

Search completed: April 24, 2001, 15:14:48
Job time: 70 sec

OM of: US-09-367-009-3 to: GenEmbl.* out_format : pfs
 Date: Apr 28, 2001 4:44 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
 -Q/cgn2.1/USPTO.spool/US09367009/runat_24042001_151806_19020/app_query.fasta_1.131
 -DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -OGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blomsu62 -TRANS=human40.cdi
 -LIST=45 -MODE=LOCAL -OUTPWT=pfs -NORM=ext -MINLEN=0
 -ALIGN=15 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR
 -MAXLEN=2000000000 -USER=US09367009_SCGN1.1.4402 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-367-009-3
 Query length: 74
 Database: GenEmbl.*
 Database sequences: 1283235
 Database length: -1216004940
 Search time (sec): 1355.860000

Score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
b.pr8:HSA224173	+	390.00	900.82	5.8e-42	496	AJ224173 Homo sapiens mRNA for
b.pat1:AX013093	+	390.00	900.43	6.1e-42	517	AX013093 Sequence 33 from Patent
b.pat1:AF071219	+	390.00	900.43	6.1e-42	517	AF071219 Homo sapiens mammaglob
b.pat1:AF074200	+	324.50	692.80	2.3e-40	17281	AF074200 Homo sapiens chromo
b.pat1:AF002793	+	324.50	692.35	2.4e-40	180676	AF002793 Homo sapiens chromo
b.pat1:AF003023	+	255.00	535.26	1.3e-21	104768	AF003023 Homo sapiens chromo
b.pat1:AF08618	+	218.00	499.15	1.4e-19	506	AF08618 Oryctolagus cuniculus
b.pat1:AR080288	+	208.00	477.96	2.1e-18	403	AR080288 Sequence 5 from patent
b.pat1:AR095406	+	208.00	477.96	2.1e-18	403	AR095406 Sequence 5 from patent
b.pat1:AR095406	+	208.00	477.96	2.1e-18	403	AR095406 Sequence 5 from patent
b.pat1:AR080288	+	208.00	475.87	2.7e-18	503	AR080288 Sequence 1 from patent
b.pat1:AR095403	+	208.00	475.87	2.7e-18	503	AR095403 Sequence 1 from patent
b.pat1:AR095403	+	208.00	475.87	2.7e-18	503	AR095403 Sequence 1 from patent
b.pat1:AF015224	+	180.00	390.50	1.6e-13	4206	AF015224 Homo sapiens mammaglob
b.pat1:AC004127	+	180.00	355.62	1.4e-11	170215	AC004127 Homo sapiens chromo
b.pat1:AF074200	+	175.00	399.97	4.6e-14	17281	AF074200 Homo sapiens chromo
b.pat1:AF08619	+	169.00	385.84	2.8e-13	452	AF08619 Oryctolagus cuniculus
b.pat1:AF08620	+	164.00	373.98	1.3e-12	461	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	163.00	370.72	2.0e-12	509	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	157.00	357.73	1.0e-11	457	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	156.50	330.74	3.3e-10	7072	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	152.00	353.47	1.8e-11	208	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	148.00	344.14	6.0e-11	208	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	147.00	333.98	2.2e-10	477	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	139.00	315.61	2.3e-09	462	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	78.50	122.81	126.61	110000	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	76.50	129.05	56.92	34595	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	72.50	122.29	135.44	26315	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	71.00	141.02	12.26	2487	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.50	103.13	1.6e-03	122379	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.50	100.51	2.2e-03	161595	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.50	100.02	2.4e-03	170279	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.50	99.56	2.5e-03	178818	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.00	130.60	46.64	5864	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.00	105.27	1.2e-03	86188	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.00	104.14	1.4e-03	97220	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.00	104.12	1.4e-03	97346	AF08620 Oryctolagus cuniculus
b.pat1:AF08620	+	70.00	103.10	1.6e-03	108469	AF08620 Oryctolagus cuniculus

gb_htg99:AC024034 + 70.00 100.31 2.3e+03 145961 | AC024034 Homo sapiens chr
 gb_htg20:AL159154 + 70.00 98.69 2.8e+03 173361 | AL159154 Homo sapiens chr
 gb_htg7:AC021123 + 70.00 97.63 3.2e+03 193802 | AC021123 Homo sapiens chr
 gb_ba3:RPXX04 + 70.00 95.72 4.1e+03 237523 | AJ235273 Rickettsia prow

seq_name: gb_pr8:HSA224173

seq_documentation_block:
 LOCUS HSA224173 496 bp mRNA PRI 25-MAR-1999
 DEFINITION Homo sapiens mRNA for lipophilin C.
 ACCESSION AJ224173
 VERSION AJ224173.1 GI:4107232
 KEYWORDS lipophilin C.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 496)
 AUTHORS Zhao,C., Nguyen,T., Yusifov,T., Glasgow,B.J. and Lehrer,R.I.
 TITLE Lipophilins: human peptides homologous to rat prostatein
 JOURNAL Biochem. Biophys. Res. Commun. 256 (1), 147-155 (1999)
 MEDLINE 99167354

REFERENCE 2 (bases 1 to 496)

AUTHORS Zhao,C.
 TITLE Direct Submission
 JOURNAL Submitted (06-FEB-1998) Zhao C., UCLA Dept. of Medicine, CHS 37055,
 Los Angeles, CA 90095, USA

FEATURES
 Location/Qualifiers
 source 1..496

/organism="Homo sapiens"
 /db_xref="taxon:9606"

/codon_start=1

/product="lipophilin C"

/protein_id="CAA11865.1"

/db_xref="GI:4107232"

/translation="MKLLMVLAAALLHCYADSGCKLLEDVETKINSISIPYKE
 LIQFIDSDAAAEAKGFKQFLNQSHRTLANFLGMHTYDTSIWNKSN"

mat_peptide 95..325

/product="lipophilin C"

BASE COUNT 141 a 119 c 102 g 134 t

ORIGIN

alignment_scores:

Quality: 390.00 Length: 74
 Ratio: 5.270 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:

US-09-367-009-3 x HSA224173

Align seg 1/1 to: HSA224173 from: 1 to: 496

1 AspSerGlyCysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 95 GATTCTGGCTGCAAACTCTGGAGGACATGTTGAAAGACCATCAATC 144
 |||||
 17 rAspIleSerIleProGluTyrLysGluLeuLeuGluGluPheIleAsp 34
 |||||
 145 CGACATATATATACCTGCAATACAAAGAGCTCTTCAGAGTTCATAGACA 194
 |||||
 34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlyCysPheLeuAsn 50
 |||||
 195 GTGATGCCGCTGCAGAGGCTATGGGAAATTCAGGAGTGTTCCTCAAC 244
 |||||
 51 GlnSerHisArgThrLeuLysAsnGlyLeuMetMetHisThrValty 67
 |||||
 245 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATGATGATGATGTA 294
 |||||
 67 rAspSerIleTyrCysAsnLeu 74
 |||||
 295 CGACAGCATTTGTGTAAATG 316

```

seq_name: gb_pat1:AX013093

seq_documentation_block:
LOCUS      AX013093      517 bp      DNA      PAT      07-SEP-2000
DEFINITION Sequence 33 from Patent WO9954461.
ACCESSION  AX013093
VERSION     AX013093.1  GI:10040259
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 517)
AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
          Pillarsky,C.
TITLE     Human nucleic acid sequences of endometrium tumour tissue
JOURNAL   Patent: WO 9954461-A 33 28-OCT-1999;
          SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
          BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
          (DE); PILLARSKY CHRISTIAN (DE)
FEATURES   Location/Qualifiers
            source          1..517
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="11"
                        /map="11q13"
                        /note="member of the uteroglobin family"
            CDS             65..352
                        /codon_start=1
                        /product="mammaglobin B precursor"
                        /protein_id="AAC79996.1"
                        /db_xref="GI:3288868"
                        /translation="MKLLMVLMLAALLHCYADSGCKLLEDVMKTKTINSIDISPEYKE
                        LLOEFIDSDAAAEAMGKFKQCFLNQSHRTLNKFLGLMHTVYDSIWNKNSN"
BASE COUNT  156 a 121 c 106 g 134 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
25
```

The sequence of Homo sapiens clone
 Unpublished
 REFERENCE 2 (bases 1 to 172281)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On Aug 11, 2000 this sequence version replaced gi:9247151.

Align seg 1/1 to: AC074200 from: 1 to: 172281
 1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 49571 GATTCGCTGCAAACTCTCTGGAGGACATGGTTGAAAAGACCATCAATC 49620
 17 rAspIleSerIleProGlyTrpLysGluLeuLeuGlnGluPheIleAspS 34
 |||||
 49621 CGACATATCTATACCTGAATACAAAGAGCTTCTCAAGAGTTTCATAGACA 49670
 34 eRAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
 |||||
 49671 GTGATCCGCTGCAGAGCTATGGGAATTCGAAGCAGTGTTCCTCAAC 49720
 51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
 |||||
 49721 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATG..... 49759
 67 rAspSerIleTrp 71
 |||||
 49760GTAATTGG 49768
 seq_name: gb_hg24:AP002793
 seq_documentation_block:
 LOCUS AP002793 180676 bp DNA HTG 25-OCT-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-691L4 map 11q, WORKING DRAFT
 SEQUENCE, 14 unordered pieces.
 ACCESSION AP002793
 VERSION AP002793.3 GI:11022730
 KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
 SOURCE Homo sapiens DNA, clone: RP11-691L4.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 180676)
 AUTHORS Hattori,M., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Published Only in DataBase (2000) In press
 JOURNAL
 REFERENCE 2 (bases 1 to 180676)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 COMMENT On Oct 26, 2000 this sequence version replaced gi:10801120.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gscc.riken.go.jp
 ----- Project Information
 Center project name: Humdraft11
 Center clone name: RP11-691L4
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 175391 bases at least Q40
 Consensus quality: 177563 bases at least Q30
 Consensus quality: 178750 bases at least Q20
 Insert size: 179376; sum-of-contigs
 Quality coverage: 8.29x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 172281)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On Aug 11, 2000 this sequence version replaced gi:9247151.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site:http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0703H08
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 166314 bases at least Q40
 Consensus quality: 168319 bases at least Q30
 Consensus quality: 169216 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 171481; sum-of-contigs
 Quality coverage: 4.93 in Q20 bases; agarose-fp
 Quality coverage: 5.45 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. * as soon as it is available and the accession number will be preserved.
 * 1 2192: contig of 2192 bp in length
 * 2193 2292: gap of unknown length
 * 2293 4860: contig of 2568 bp in length
 * 4861 4961: gap of unknown length
 * 4962 18883: contig of 13923 bp in length
 * 18884 33366: contig of 14383 bp in length
 * 33367 33466: gap of unknown length
 * 33467 52219: contig of 18753 bp in length
 * 52220 52320: gap of unknown length
 * 52320 69757: gap of unknown length
 * 69758 94565: contig of 24708 bp in length
 * 94566 120988: gap of unknown length
 * 120989 121089: contig of 26323 bp in length
 * 121089 172281: contig of 51193 bp in length.
 ----- Location/Qualifiers -----
 1..172281
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-703H8"
 BASE COUNT 49475 a 37393 c 36287 g 48325 t 801 others
 ORIGIN

alignment_scores:
 Quality: 324.50 Length: 71
 Ratio: 4.992 Gaps: 1
 Percent Similarity: 91.549 Percent Identity: 91.549
 alignment_block:
 US-09-367-009-3 x AC074200

as soon as it is available and the accession number will be preserved

```

1 56642 contig of 56642 bp in length
56743 92253 contig of 35511 bp in length
92354 116265 contig of 23912 bp in length
116366 128366 contig of 12001 bp in length
128467 138214 contig of 9748 bp in length
138315 147234 contig of 8920 bp in length
147335 155435 contig of 8101 bp in length
155536 162703 contig of 7168 bp in length
162804 167910 contig of 5107 bp in length
168011 171323 contig of 3313 bp in length
171424 174108 contig of 2685 bp in length
174209 176988 contig of 2780 bp in length
177089 178989 contig of 1901 bp in length
179090 180676 contig of 1587 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

```

1 56642: contig of 56642 bp in length
56643 56742: gap of 100 bp
56743 92253: contig of 35511 bp in length
92254 92353: gap of 100 bp
92354 116265: contig of 23912 bp in length
116266 116366: gap of 100 bp
116366 128366: contig of 12001 bp in length
128367 128466: gap of 100 bp
128467 138214: contig of 9748 bp in length
138215 138314: gap of 100 bp
138315 147234: contig of 8920 bp in length
147235 147334: gap of 100 bp
147335 155435: contig of 8101 bp in length
155436 155535: gap of 100 bp
155536 162703: contig of 7168 bp in length
162704 162803: gap of 100 bp
162804 167910: contig of 5107 bp in length
167911 168010: gap of 100 bp
168011 171323: contig of 3313 bp in length
171324 171423: gap of 100 bp
171424 174108: contig of 2685 bp in length
174109 174208: gap of 100 bp
174209 176988: contig of 2780 bp in length
176989 177088: gap of 100 bp
177089 178989: contig of 1901 bp in length
178990 179089: gap of 100 bp
179090 180676: contig of 1587 bp in length.

```

FEATURES

source

```

1..180676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-691L4"

```

```

misc_feature
1..56642
/note="assembly_fragment"
56743..92253
/note="assembly_fragment"
92354..116265
/note="assembly_fragment"
116366..128366
/note="assembly_fragment"
128467..138214
/note="assembly_fragment"
138315..147234
/note="assembly_fragment"
147335..155435
/note="assembly_fragment"
155536..162703

```

```

misc_feature
162804..167910
/note="assembly_fragment clone_end:SP6 vector_side:left"
168011..171323
/note="assembly_fragment"
171424..174108
/note="assembly_fragment clone_end:T7 vector_side:left"
174209..176988
/note="assembly_fragment"
177089..178989
/note="assembly_fragment"
179090..180676
/note="assembly_fragment"
BASE COUNT 49650 a 39805 c 39872 g 50049 t 1300 others
ORIGIN

```

```

alignment_scores:
Quality: 324.50 Length: 71
Ratio: 4.992 Gaps: 1
Percent Similarity: 91.549 Percent Identity: 91.549

```

alignment_block:

US-09-367-009-3 x AP002793/rev ..

Align seg 1/1 to reverse of: AP002793 from: 1 to: 180676

```

1 ASPSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
92220 GATTCTGCTGCAAACTCCTGGAGGACATGGTTGAAAGACCATCAATTC 92171
|||||
17 rASpSerIleProGluTyrLysGluLeuGluGlnGluPheIleAsps 34
|||||
92170 CGACATATCTATACCTCAATACAAAGAGCTTCTCAAGAGTTCATAGACA 92121
|||||
34 erASpAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
92120 GTGATGCGCTGCAGAGGCTATGGGAAATTCAAAGCAGTGTTCCTCAAC 92071
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
92070 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATG..... 92032
|||||
67 rASpSerIleTirp 71
|||||
92031 ....GTAATTGG 92023

```

seq_name: gb_htgl:AC003023

seq_documentation_block:

```

LOCUS AC003023 104768 bp DNA HTG 21-OCT-1997
DEFINITION Homo sapiens chromosome 11 clone PDJ363p2, *** SEQUENCING IN
PROGRESS ***, 22 unordered pieces.

```

AC003023

VERSION AC003023.1 GI:2554969

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 104768)

Evans, G.A., Athanasiou, M., Hahner, L., Osborne-Lawrence, S., Franklin, T.L., Federova, N., English, C., Hinson-Cooper, S., Dunn, J., McFarland, J., Davie, J., Ward, T., Card, P., Patel, P., Gordon, M., Newton, J., Valenzuela, D., Schageman, J., Harris, J., Gotway, G., Syed, M., Kupfer, K., Schilling, P., Gee, V., Basit, M., Brignac, S., Grant, O., Bumeister, R., O'Brien, K., Garner, H.R. and Schultz, R.

Template Unpublished

JOURNAL

2 (bases 1 to 104768)

Evans, G.A., Athanasiou, M., Bradbury, P., Brignac, S., Bumeister, R., Davis, C., English, C., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S.,

60528 CAGTCCCAATAAAACTCTGGAAAACCTTTGGACTG 60562

seq_documentation_block	LOCUS	AR080288	403 bp	DNA	PAT

27-04-2004

Copied from PCT009367009 on 27-04-2009

268 ACGGATGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGTCTTTGTGATTGA 336

seq_name: gb_pat1:AR080285

seq_documentation_block:
LOCUS AR080285 503 bp DNA 31-AUG-2000
DEFINITION Sequence 1 from patent US 5968754.
ACCESSION AR080285
VERSION AR080285.1 GI:10007020
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson, M.A. and Fleming, T.P.
TITLE Mammaglobin, a mammary-specific breast cancer protein
JOURNAL Patent: US 5968754-A 1 19-OCT-1999;
FEATURES
Location/Qualifiers
source
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x AR080285
Align seg 1/1 to: AR080285 from: 1 to: 503
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATTGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCAATATATGA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGGATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317

seq_name: gb_pat1:AR095403
seq_documentation_block:
LOCUS AR095403 503 bp DNA 08-SEP-2000
DEFINITION Sequence 1 from patent US 6004756.
ACCESSION AR095403
VERSION AR095403.1 GI:10023248
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson, M.A. and Fleming, T.P.
TITLE Method for detecting the presence of breast cancer by detecting an
increase in mammaglobin mRNA expression
Patent: US 6004756-A 1 21-DEC-1999;
Location/Qualifiers
source
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x I65735
Align seg 1/1 to: I65735 from: 1 to: 503
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATTGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCAATATATGA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGGATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGTCTTTGTGATTGA 336

increase in mammaglobin mRNA expression
Patent: US 6004756-A 1 21-DEC-1999;
Location/Qualifiers
source
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x AR095403
Align seg 1/1 to: AR095403 from: 1 to: 503
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATTGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCAATATATGA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGGATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGTCTTTGTGATTGA 336

seq_name: gb_pat2:I65735
seq_documentation_block:
LOCUS I65735 503 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 5668267.
ACCESSION I65735
VERSION I65735.1 GI:2482305
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 503)
AUTHORS Watson, M.A. and Fleming, T.P.
TITLE Polynucleotides encoding mammaglobin, a mammary-specific breast
cancer protein
Patent: US 5668267-A 1 16-SEP-1997;
Location/Qualifiers
source
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x I65735
Align seg 1/1 to: I65735 from: 1 to: 503
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGTGCCTTATTGAGAAATGTTTCCAGACAAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuGluGlnPheIleAspSera 35
168 AGTGCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCAATATATGA 217
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATTCAGAGGATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGTCTTTGTGATTGA 336

```

118 TCTGGTCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 p1eSerIleProGluTyLysGluLeuGlnGluPheIleAspSera 35
168 AGTGTCTAAGACTGAATACAAAGACTTCTTCAAGAGTTCATAGACGACA 217
35 sPaLaLaLaGluAlaMetGlyLysPheGlyGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyra 68
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTTA 336
seq_name: gb_pr9:HSU33147
seq_documentation_block:
LOCUS HSU33147 503 bp mRNA PRI 22-FEB-1996
DEFINITION Human mammaglobin mRNA, complete cds.
ACCESSION U33147
VERSION U33147.1 GI:1199595
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Mamaglobin, a mammary-specific member of the uteroglobin gene
family, is overexpressed in human breast cancer
JOURNAL Cancer Res. 56 (4), 860-865 (1996)
MEDLINE 96223698
REFERENCE
2 (bases 1 to 503)
AUTHORS Watson,M.A.
DIRECT SUBMISSION
Submitted (03-AUG-1995) Mark A. Watson, Washington Univ. School of
Medicine, Dept. Ophthalmology and Visual Sciences, Dept. Genetics,
Div. Laboratory Medicine, Box 8118, 660 S. Euclid Avenue, St.
Louis, MO 63110, USA
FEATURES
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"
61..342
/Note="similar to uteroglobin: SwissProt Accession Number
P02779 and rat prostatic steroid binding protein subunit
C3: SwissProt Accession Number P02780"
/codon_start=1
/product="mammaglobin"
/protein_id="AAC50375.1"
/db_xref="GI:1199596"
/translation="MKLLMVLMAALSOHCYAGSGCPLEENVISKTNPOVSKTEYKE
LQEFIDNATNNAIDELKECFNLQDTLSNVVEFMQLIYDSSICDLF"
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x HSU33147
Align seg 1/1 to: HSU33147 from: 1 to: 503
118 TCTGGTCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 p1eSerIleProGluTyLysGluLeuGlnGluPheIleAspSera 35
168 AGTGTCTAAGACTGAATACAAAGACTTCTTCAAGAGTTCATAGACGACA 217
35 sPaLaLaLaGluAlaMetGlyLysPheGlyGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyra 68
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTTA 336
seq_name: gb_pr9:HSU33147
seq_documentation_block:
LOCUS HSU33147 503 bp mRNA PRI 22-FEB-1996
DEFINITION Human mammaglobin mRNA, complete cds.
ACCESSION U33147
VERSION U33147.1 GI:1199595
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 503)
AUTHORS Watson,M.A. and Fleming,T.P.
TITLE Mamaglobin, a mammary-specific member of the uteroglobin gene
family, is overexpressed in human breast cancer
JOURNAL Cancer Res. 56 (4), 860-865 (1996)
MEDLINE 96223698
REFERENCE
2 (bases 1 to 503)
AUTHORS Watson,M.A.
DIRECT SUBMISSION
Submitted (03-AUG-1995) Mark A. Watson, Washington Univ. School of
Medicine, Dept. Ophthalmology and Visual Sciences, Dept. Genetics,
Div. Laboratory Medicine, Box 8118, 660 S. Euclid Avenue, St.
Louis, MO 63110, USA
FEATURES
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"
61..342
/Note="similar to uteroglobin: SwissProt Accession Number
P02779 and rat prostatic steroid binding protein subunit
C3: SwissProt Accession Number P02780"
/codon_start=1
/product="mammaglobin"
/protein_id="AAC50375.1"
/db_xref="GI:1199596"
/translation="MKLLMVLMAALSOHCYAGSGCPLEENVISKTNPOVSKTEYKE
LQEFIDNATNNAIDELKECFNLQDTLSNVVEFMQLIYDSSICDLF"
BASE COUNT 146 a 118 c 97 g 142 t
ORIGIN
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x HSU33147
Align seg 1/1 to: HSU33147 from: 1 to: 503
118 TCTGGTCCCTTATTGGAGATGTGATTTCCAAGACAATCAATCCACA 167
18 p1eSerIleProGluTyLysGluLeuGlnGluPheIleAspSera 35
168 AGTGTCTAAGACTGAATACAAAGACTTCTTCAAGAGTTCATAGACGACA 217
35 sPaLaLaLaGluAlaMetGlyLysPheGlyGlnCysPheLeuAsnGln 51
218 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyra 68
268 ACGGATGAACACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74
318 CAGCAGCTCTTTGTGATTTA 336
seq_name: gb_pr4:AF015224
seq_documentation_block:
LOCUS AF015224 4206 bp DNA PRI 25-FEB-1998
DEFINITION Homo sapiens mammaglobin gene, complete cds.
ACCESSION AF015224
VERSION AF015224.1 GI:2911295
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4206)
AUTHORS Watson,M.A., Darrow,C., Zimonjic,D.B., Popescu,N.C. and
Fleming,T.P.
TITLE Structure and transcriptional regulation of the human mammaglobin
gene, a breast cancer associated member of the uteroglobin gene
family localized to chromosome 11q13
JOURNAL Oncogene 16 (6), 817-824 (1998)
MEDLINE 98147371
REFERENCE
2 (bases 1 to 4206)
AUTHORS Watson,M.A., Fleming,T.P. and Darrow,C.
DIRECT SUBMISSION
Submitted (19-JUL-1997) Department Pathology / Box 8118, Washington
University School of Medicine, 660 S. Euclid Ave., St. Louis, MO
63110, USA
FEATURES
Location/Qualifiers
1..4206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="11"
/map="11q12.3-13.1"
10..232
/Note="contains ERE-like element"
/rept_family="Alu"
Join(992..1110,1713..1900,3789..3974)
/product="mammaglobin"
/protein_id="AAC39608.1"
/db_xref="GI:2911296"
/translation="MKLLMVLMAALSOHCYAGSGCPLEENVISKTNPOVSKTEYKE
LQEFIDNATNNAIDELKECFNLQDTLSNVVEFMQLIYDSSICDLF"
BASE COUNT 1319 a 947 c 836 g 1103 t 1 others
ORIGIN
alignment_scores:
Quality: 180.00 Length: 68
Ratio: 3.158 Gaps: 0

```

```
Percent Similarity: 83.824 Percent Identity: 51.471
alignment_block:
US-09-367-009-3 x AF015224 ..
Align seg 1/1 to: AF015224 from: 1 to: 4206

2 SerGlyCysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1715 TCTGGCTGCCCTTATTGGAGATGTGATTCCAAGACAATCAATCCACA 1764

18 pIleSerIleProGluTyrLysGluLeuGluGluPheIleAspSera 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1765 AGTGTCTAAGACTGAATCAAGAACTTCTCAAGAGTTCATAGACGACA 1814

35 spAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1815 ATGCCACTACAAATGCCATAGATGAATTGAAGGAATGTTTCTTAACCAA 1864

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1865 ACGGATGAAACTCTGAGCAATCTTGAGGTGTTTATGGTAATTTCATTTC 1914

68 pSer 69
|||
1915 TTCC 1918
```

OM of: US-09-367-009-3 to: N_Geneseq_0401.* out_format : pfs
Date: Apr 28, 2001 4:48 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framed_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09367009/runat_24042001_151807_19057/app_query.fasta_1.131
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09367009.ecgn1_1.174 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-367-009-3
Query length: 74
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890851
Search time (sec): 112.140000

core_list:	Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
SIDS2/gcgdata/geneseq/NA2000.DAT:243008 +	SIDS2/gcgdata/geneseq/NA2000.DAT:243008 +	390.00	1007.39	3.6e-48	473	
SIDS2/gcgdata/geneseq/NA1997.DAT:794832 +	SIDS2/gcgdata/geneseq/NA1997.DAT:794832 +	390.00	1007.32	3.6e-48	476	
SIDS2/gcgdata/geneseq/NA2000.DAT:A59730 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A59730 +	390.00	1007.32	3.6e-48	476	
SIDS2/gcgdata/geneseq/NA2000.DAT:A09104 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A09104 +	390.00	1007.32	3.6e-48	476	
SIDS2/gcgdata/geneseq/NA1999.DAT:X36138 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X36138 +	390.00	1007.12	3.7e-48	485	
SIDS2/gcgdata/geneseq/NA1999.DAT:X36138 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X36138 +	390.00	1006.45	4.0e-48	517	
SIDS2/gcgdata/geneseq/NA2000.DAT:A09118 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A09118 +	390.00	1006.45	4.0e-48	517	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	208.00	534.76	7.6e-22	279	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	208.00	530.87	1.2e-21	403	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	208.00	528.69	1.6e-21	495	
SIDS2/gcgdata/geneseq/NA1998.DAT:A11580 +	SIDS2/gcgdata/geneseq/NA1998.DAT:A11580 +	208.00	528.52	1.7e-21	503	
SIDS2/gcgdata/geneseq/NA1997.DAT:T50925 +	SIDS2/gcgdata/geneseq/NA1997.DAT:T50925 +	208.00	528.52	1.7e-21	503	
SIDS2/gcgdata/geneseq/NA1998.DAT:A17905 +	SIDS2/gcgdata/geneseq/NA1998.DAT:A17905 +	208.00	528.52	1.7e-21	503	
SIDS2/gcgdata/geneseq/NA1999.DAT:X26986 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X26986 +	208.00	528.52	1.7e-21	503	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	208.00	528.52	1.7e-21	503	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12632 +	208.00	528.36	1.7e-21	511	
SIDS2/gcgdata/geneseq/NA1998.DAT:F17694 +	SIDS2/gcgdata/geneseq/NA1998.DAT:F17694 +	208.00	527.62	1.9e-21	548	
SIDS2/gcgdata/geneseq/NA2001.DAT:F17909 +	SIDS2/gcgdata/geneseq/NA2001.DAT:F17909 +	185.00	471.75	2.4e-18	356	
SIDS2/gcgdata/geneseq/NA2000.DAT:F17729 +	SIDS2/gcgdata/geneseq/NA2000.DAT:F17729 +	67.00	118.67	113.56	20844	
SIDS2/gcgdata/geneseq/NA1998.DAT:A81460 +	SIDS2/gcgdata/geneseq/NA1998.DAT:A81460 +	66.00	147.75	2.72	1038	
SIDS2/gcgdata/geneseq/NA1998.DAT:A44253 +	SIDS2/gcgdata/geneseq/NA1998.DAT:A44253 +	64.50	152.37	1.51	462	
SIDS2/gcgdata/geneseq/NA1999.DAT:X99624 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X99624 +	64.50	147.19	2.93	754	
SIDS2/gcgdata/geneseq/NA2000.DAT:A01735 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A01735 +	64.50	137.80	9.76	1833	
SIDS2/gcgdata/geneseq/NA2000.DAT:C49471 +	SIDS2/gcgdata/geneseq/NA2000.DAT:C49471 +	64.50	120.71	87.37	9229	
SIDS2/gcgdata/geneseq/NA1998.DAT:N80437 +	SIDS2/gcgdata/geneseq/NA1998.DAT:N80437 +	64.00	145.67	3.56	769	
SIDS2/gcgdata/geneseq/NA1999.DAT:Z17472 +	SIDS2/gcgdata/geneseq/NA1999.DAT:Z17472 +	64.00	145.67	3.56	769	
SIDS2/gcgdata/geneseq/NA1999.DAT:Z17473 +	SIDS2/gcgdata/geneseq/NA1999.DAT:Z17473 +	62.50	117.71	128.40	8144	
SIDS2/gcgdata/geneseq/NA2000.DAT:A70106 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A70106 +	62.50	116.78	144.68	8144	
SIDS2/gcgdata/geneseq/NA1999.DAT:X32482 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X32482 +	62.00	128.86	30.74	2295	
SIDS2/gcgdata/geneseq/NA2000.DAT:A70217 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A70217 +	62.00	118.62	114.23	6042	
SIDS2/gcgdata/geneseq/NA2000.DAT:A70199 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A70199 +	62.00	115.53	169.78	8093	
SIDS2/gcgdata/geneseq/NA2000.DAT:A81512 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A81512 +	62.00	75.71	2.8e+04	349980	
SIDS2/gcgdata/geneseq/NA2000.DAT:F21609 +	SIDS2/gcgdata/geneseq/NA2000.DAT:F21609 +	62.00	75.71	2.8e+04	349980	
SIDS2/gcgdata/geneseq/NA2000.DAT:F21610 +	SIDS2/gcgdata/geneseq/NA2000.DAT:F21610 +	62.00	60.77	1.8e+05	1437668	
SIDS2/gcgdata/geneseq/NA2000.DAT:A81490 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A81490 +	61.50	121.72	76.81	3982	
SIDS2/gcgdata/geneseq/NA1998.DAT:V26083 +	SIDS2/gcgdata/geneseq/NA1998.DAT:V26083 +	61.50	112.12	263.00	9870	
SIDS2/gcgdata/geneseq/NA1998.DAT:V13935 +	SIDS2/gcgdata/geneseq/NA1998.DAT:V13935 +	61.50	112.12	263.00	9870	
SIDS2/gcgdata/geneseq/NA1999.DAT:V16457 +	SIDS2/gcgdata/geneseq/NA1999.DAT:V16457 +	61.50	106.45	544.02	16870	
SIDS2/gcgdata/geneseq/NA1999.DAT:X13035 +	SIDS2/gcgdata/geneseq/NA1999.DAT:X13035 +	61.50	106.45	544.02	16870	
SIDS2/gcgdata/geneseq/NA1998.DAT:V26084 +	SIDS2/gcgdata/geneseq/NA1998.DAT:V26084 +	61.00	125.26	48.75	2515	
SIDS2/gcgdata/geneseq/NA1998.DAT:V11152 +	SIDS2/gcgdata/geneseq/NA1998.DAT:V11152 +	61.00	125.26	48.75	2515	
SIDS2/gcgdata/geneseq/NA2000.DAT:A12387 +	SIDS2/gcgdata/geneseq/NA2000.DAT:A12387 +	60.50	134.37	15.15	938	

/SIDS2/gcgdata/geneseq/NA2000.DAT:C49309 + 60.50 129.30 29.04 1516
/SIDS2/gcgdata/geneseq/NA2000.DAT:C41123 + 60.50 129.28 29.10 1518
/SIDS2/gcgdata/geneseq/NA1998.DAT:V11858 + 60.50 128.09 33.92 1700
/SIDS2/gcgdata/geneseq/NA2000.DAT:A81516 + 60.00 117.29 135.51 4169
/SIDS2/gcgdata/geneseq/NA2000.DAT:F21611 + 60.00 70.45 5.4e+04 349980

seq_name: /SIDS2/gcgdata/geneseq/NA2000.DAT:243008

seq_documentation_block:

ID_243008 standard; CDNA; 473 BP.

AC 243008;

DT 01-FEB-2000 (first entry)

DE Human 5' EST isolated from a cDNA library SEQ ID NO:767.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.

XX Homo sapiens.

OS Homo sapiens.

PN WO9953051-A2.

XX 21-OCT-1999.

PD 09-APR-1999; 99WO-IB00712.

PF 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-038446/03.

XX P-PSDB; Y65394.

XX Novel secreted protein 5' expressed sequence tag sequences used in
diagnostic, forensic, gene therapy, and chromosome mapping procedures
Claim 1; Page 569; 837pp; English.

CC Z42265 to Z43075 represent novel 5' expressed sequence tag (EST)
sequences, corresponding to human secreted proteins. Y64651 to Y65438
represent the EST-related proteins corresponding to Z42265 to Z43052.
The 5' ESTs can be used for producing secreted human gene products.
They can be used to identify and isolate 5' untranslated regions (UTRs)
and upstream regulatory regions which control the location, development
stage, rate, and quantity of protein synthesis, as well as stability of
mRNA. The ESTs are also useful as probes for chromosome mapping, and to
obtain full length cDNA clones. The ESTs can also be used in forensic
procedures to identify individuals, or in diagnostic procedures to
identify individuals having genetic diseases resulting from abnormal
gene expression. The products may also be used in gene therapy protocols.
The nucleic acids encoding signal peptides can be used for directing
extracellular secretion of a polypeptide or the insertion of a
polypeptide into a membrane, or importing a polypeptide into a cell.
The proteins encoded by the EST sequences may be useful in treating a
variety of human conditions. Secreted proteins have therapeutic value,
and the identification of new secreted proteins is valuable. Z42249 to
Z42264 and Y64644 to Y64650 represent sequences used in the
exemplification of the present invention.

XX Sequence 473 BP; 130 A; 113 C; 104 G; 123 T; 3 other;

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

This invention describes novel human nucleic acid (cDNA) sequences (A), that are highly expressed in uterine tumour tissue and which have anticancer and cytostatic activity. (A) are used (i) for recombinant expression of polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to identify agents suitable for treatment of uterine or endometrial cancer; (ii) directly for treating these forms of cancer (including expression from gene therapy vectors) and (iii) for generation of specific antibodies. (A) are identified by assembling ESTs (expressed sequence tags) from a particular tissue type before comparison of expression patterns. This allows a significantly longer fragment of the gene to be revealed, so should reduce the number of failures associated with the fact that ESTs from different libraries may represent different parts of the same unknown gene, distorting the estimated frequency of occurrence in a particular tissue. Z41981-742121 represent EST fragments obtained from a human endometrium tumour cDNA library which encode the protein sequences represented in Y59941-Y60328.

Sequence 517 BP; 156A; 121 C; 106 G; 134 T; 0 other;

```

alignment_scores:
  Quality: 390.00      Length: 74
  Ratio: 5.270        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 98.649

alignment_block:
  US-09-367-009-3 x 242013  ..

Align seg 1/1 to: 242013 from: 1 to: 517

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
  |||||
95 GATTCGGCTGCCAACTCCTCGAGGACATGGTTGAAAGACCATCAATTC 144

17 rAspIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspS 34
  |||||
145 CGACATATCTATACCTGATACAAAGAGCTTCTTCAGAGTCTCATGACA 194

34 eRAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
  |||||
195 GTGATGCCCTGCAGAGGCTATGGGGAATTCACAGCAGTGTTCCTCAAC 244

51 GlnSerHisArThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
  |||||
245 CAGTCACATAGAACTCTGAAACATTGGGACTGATGATGCATCAGTGTA 294

67 rAspSerIleIleTrpCysAsnLeu 74
  |||||
295 CGACAGCATTTGGTGAATG 316

```

seq_name: /SIDS2/gcgcdata/geneseq/geneseq/NA2000.DAT:A09118

req_documentation_block:	
D	A09118 standard; cDNA; 517 BP.
XX	
XX	A09118;
XX	
XX	
XX	10-AUG-2000 (first entry)
XX	
XX	Clone Mamm-X cDNA, encodes mamoglobin homologue.
DE	
XX	
XX	Clone Mamm-X; mamoglobin; breast cancer; cytostatic; anti-HIV;
XX	immunosuppressive; antiallergic; antineoplastic; antiinflammatory;
XX	antiarthritic; antiarteriosclerotic; vasotropic; neuroprotective;
XX	antitropic; dermatological; tranquilizer; vulnery; ss.
XX	
XX	Homo sapiens.
XX	

Key	Location/Qualifiers
CDS	65..352
.	/*tag= a

```

PN      WO200020447-A2.
XX      13-APR-2000.
XX      06-OCT-1999; 99WO-US23294.
XX      06-OCT-1998; 98US-0103195.
XX      03-OCT-1999; 99US-0103195.
XX      (CURA-) CURAGEN CORP.
XX      Shimkets RA;
XX      WPI; 2000-303741/26.
XX      P-PSDB; Y92237.
XX      Nucleic acids encoding polypeptides with syncollin-like, C-
XX      cytokine-like activity, useful for treating diseases inclu
XX      Alzheimer's and atherosclerosis
XX      Claim 28; Fig 6; 118pp; English.
XX      Clone Mamm-X encodes a polypeptide that is 100 percent iden
XX      Mammaglobin B precursor, a potential marker of breast can
XX      metastasis. The sequences are useful for treatment of dis
XX      cancer, immune disorders, autoimmune disease, acquired imm
XX      syndrome (AIDS), transplant rejection, allergy, infection
XX      pathological agent or organism, inflammatory disorders, ar
XX      haematopoietic disorder, a skin disorder, atherosclerosis,
XX      neurological disease, Alzheimer's disease, trauma, spinal
XX      and skeletal disorders.
XX      Sequence 517 BP; 144 A; 129 C; 105 G; 139 T; 0 other;

alignment_scores:
    Quality: 390.00      Length: 74
    Ratio: 5.270        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x A09118 ..

Align seg 1/1 to: A09118 from: 1 to: 517

1 ASPSerGlyCysIysLeuLeuGluAspMetValGluLyThrIleAsnSe 17
|||||
119 GATCTGGCTGCAAACTCCTGGAGGACATGGTTGAAAAGACCATCAATTC 168
|||||

17 rAspIleSerIleProGluTyrIysGluLeuLeuGluGluPheIleAspS 34
|||||
169 CGACATATCTATACCTGGAATACAAAGAGCTTCCTCAAGAGTTCATAGACA 218
|||||

34 erAspAlaAlaGluAlaMetGlyLySPheLysGlnCysPheLeuAsn 50
|||||
219 GTGATCCGCGTCGAGAGGCTATGGGGAATTCNAGCAGTGTCTTCCTCAC 268
|||||

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTr 67
|||||
269 CAGTCACATAGAACTCTGAAAACATTGTGACATGATGCATGCATACAGTGA 318
|||||

67 rAspSerIleThrCysAsnLeu 74
|||||
319 CGACAGCATTTGGTGTAAATG 340
|||||

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A12642

seq_documentation_block:
ID AL2642 standard; cDNA; 279 BP.
XX
AC AL2642:

DT 25-JUL-2000 (first entry)

Copied from PCT009367009 on 27-04-2004

Fragment of cDNA encoding the mammary-specific protein mammaglobin.
Human; mammaglobin; mammary gland; breast cancer; endometrial cancer; ss.
Homo sapiens.
WO200018783-A1.
06-APR-2000.
29-SEP-1999; 99WO-US22616.
29-SEP-1998; 98US-0162622.
(UNIW) UNIV WASHINGTON.
Watson MA, Fleming TP;
WPI; 2000-293105/25.
Methods for detecting breast cancer, comprising detecting elevated concentrations of a mammaglobin polypeptide, using an antibody, or detecting elevated concentrations of the mRNA encoding the polypeptide, using oligonucleotides -
Claim 3; Page 67; 7lpp; English.
The present sequence represents a fragment of cDNA encoding the human mammary-specific secreted protein mammaglobin. Mammaglobin expression is restricted to the mammary gland. Dysregulation of the mammaglobin gene occurs early and frequently in breast cancer. The specification describes a method for detecting the presence of breast cancer in a patient, comprising detecting an elevated concentration of mRNA encoding a mammaglobin polypeptide. The methods are useful for detecting the presence of breast and endometrial cancer.
Sequence 279 BP; 83 A; 58 C; 58 G; 80 T; 0 other;
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x A12642 ..
Align seg 1/1 to: A12642 from: 1 to: 279
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
58 TCTGGCTGCCCTTATTGGAGAACTGTTTCCAGACAATCAATCCACA 107
18 pIleSerIleProGluTyrLysGluLeuLeuGluPheIleAspSera 35
108 AGTGCTTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 157
35 sPAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
158 ATGCCACTACAATGCCATAGATGAATTCAGAGAACTGTTTCTTAACCAA 207
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
208 ACGGATGAAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 257
68 pSerIleTrpCysAsnLeu 74
258 CAGCAGCTCTTCTGTGATTTA 276
seq_name: /SID52/cgdata/geneseq/geneseq/NA2000.DAT:A12635
seq_documentation_block:

Al2635 standard; cDNA; 403 BP.
Al2635;
25-JUL-2000 (first entry)
PCR amplified 403 bp mammaglobin cDNA fragment.
Human; mammaglobin; mammary gland; breast cancer; endometrial cancer; ss.
Homo sapiens.
WO200018783-A1.
06-APR-2000.
29-SEP-1999; 99WO-US22616.
29-SEP-1998; 98US-0162622.
(UNIW) UNIV WASHINGTON.
Watson MA, Fleming TP;
WPI; 2000-293105/25.
Methods for detecting breast cancer, comprising detecting elevated concentrations of a mammaglobin polypeptide, using an antibody, or detecting elevated concentrations of the mRNA encoding the polypeptide, using oligonucleotides -
Example 1; Page 65; 7lpp; English.
The present sequence represents a PCR amplified fragment of cDNA encoding the human mammary-specific secreted protein mammaglobin. Mammaglobin expression is restricted to the mammary gland. Dysregulation of the mammaglobin gene occurs early and frequently in breast cancer. The specification describes a method for detecting the presence of breast cancer in a patient, comprising detecting an elevated concentration of mRNA encoding a mammaglobin polypeptide. The methods are useful for detecting the presence of breast and endometrial cancer.
Sequence 403 BP; 115 A; 97 C; 86 G; 105 T; 0 other;
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
alignment_block:
US-09-367-009-3 x A12635 ..
Align seg 1/1 to: A12635 from: 1 to: 403
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
118 TCTGGCTGCCCTTATTGGAGAACTGTTTCCAGACAATCAATCCACA 167
18 pIleSerIleProGluTyrLysGluLeuLeuGluPheIleAspSera 35
168 AGTGCTTAAGACTGAATACAAAGAACTTCTCAAGAGTTTCATAGACGACA 217
35 sPAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
218 ATGCCACTACAATGCCATAGATGAATTCAGAGAACTGTTTCTTAACCAA 267
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
268 ACGGATGAAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATGA 317
68 pSerIleTrpCysAsnLeu 74

```

|||||
318 CAGCAGTCTTTGTGATTTA 336

seq_name: /SIDS2/gcgdata/geneseq/NA1998.DAT:V41580
seq_documentation_block:
ID V41580 standard; CDNA; 495 BP.
XX AC V41580;
XX
XX 12-OCT-1998 (first entry)
XX
XX Nucleotide sequence of the human steroid binding protein C2.
DE
XX
XX Human steroid-binding protein C2; hSBP2; hSBP1; breast cancer; probe;
KW gene therapy vector; ribozyme; probe; hybridisation; amplification;
KW antibody; immunoassay; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 46..327
XX /tag= a
XX /product= "human steroid binding protein C2"
XX
XX WO9821331-AL.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-US20674.
XX
XX 12-NOV-1996; 96US-0747547.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Akerblom IE, Goli SK, Hawkins PR, Hillman JL, Murry LE;
XX WPI; 1998-297935/26.
XX P-PSDB; W59777.
XX
XX New human steroid binding proteins C1 and C2 - useful for, e.g.
XX diagnosis, monitoring and treating breast cancer, and for drug
XX screening
XX
XX Claim 14; Fig 2; 70pp; English.
XX
XX This is the nucleotide sequence of the human steroid-binding protein
XX C2 (hSBP2) used in the method of the invention for the diagnosis,
XX monitoring and treatment of breast cancer. HSBP1 and hSBP2 are useful
XX as markers for breast cancer, i.e. measuring levels of hSBP1 and hSBP2
XX used for diagnosis or monitoring the disease, to identify subjects
XX at risk and to discriminate between different forms of cancer for
XX selection of appropriate therapies. They may also be used for drug
XX screening. Nucleic acids encoding hSBP1 and hSBP2 can be used in gene
XX therapy vectors to overexpress the steroid-binding proteins, preventing
XX binding of steroids, or antisense sequences, ribozymes. Their nucleic
XX acids can also be used for the diagnosis and monitoring (by quantifying
XX expression of hSBP), as source of probes for hybridisation and
XX amplification of genomic or related sequences for studying regulation of
XX gene function and for mapping the genomic sequence. Antibodies are used
XX as diagnostic reagents in standard immunoassays for hSBP.
XX
XX Sequence 495 BP; 148 A; 114 C; 94 G; 139 T; 0 other;

```

```

alignment_scores:
  Quality: 208.00      Length: 73
  Ratio: 3.355        Gaps: 0
  Percent Similarity: 84.932  Percent Identity: 53.425

```

```

alignment_block:
US-09-367-009-3 x V41580

```

```

Align seg 1/1 to: V41580 from: 1 to: 495

2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
103 TCTGGCTGCCCTTATTGGAGATGTGATTCCAAGACAAATCAATCCACA 152
18 pIleSerIleProGluTyrLysGluLeuLeuGluPheIleAspSerA 35
153 AGTGTCTAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGACGACA 202
35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
203 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 252
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
253 ACGGATGAACCTCTGAGCAATGTTGAGTGTGTTATGCAATTAATATGA 302
68 pSerIleTrpCysAsnLeu 74
303 CAGCAGTCTTTGTGATTTA 321

seq_name: /SIDS2/gcgdata/geneseq/NA1997.DAT:T50925
seq_documentation_block:
ID T50925 standard; CDNA; 503 BP.
XX
XX AC T50925;
XX
XX 12-AUG-1997 (first entry)
XX
XX DE cDNA encoding mammary-specific secretory protein, mammaglobin.
XX
XX KW mammaglobin; mammary-secretory protein; breast cancer; detection;
XX neoplastic disease; diagnosis; ss.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 5'UTR 1..60
XX /tag= a
XX CDS 61..342
XX /tag= b
XX /product= mammaglobin
XX sig_peptide 61..117
XX /tag= c
XX mat_peptide 118..339
XX /tag= d
XX 3'UTR 343..503
XX /tag= e
XX polyA_signal 489..494
XX /tag= f
XX misc_feature 1..403
XX /tag= g
XX /note= "403 bp fragment isolated by RACE PCR"
XX misc_feature 298..503
XX /tag= h
XX /note= "DEST002"
XX
XX WO9638463-Al.
XX
XX 05-DEC-1996.
XX
XX 31-MAY-1996; 96WO-US08235.
XX
XX 31-MAY-1995; 95US-0455896.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Fleming TP, Watson MA;
XX WPI; 1997-034299/03.
XX P-PSDB; W10179.
XX
XX DR

```



```

318 CAGCACTCTTTGTGATTGA 336
seq_name: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT:X26966
seq_documentation_block:
ID_X26966 standard; CDNA to mRNA; 503 BP.
XX
AC_X26966;
XX
25-JUN-1999 (first entry)
XX
DE cDNA encoding mamaglobin, a mammary specific protein.
XX
KW Human; mammary-specific protein; mamaglobin; antigen; vaccine;
mamaglobin-expressing cancer; breast cancer;
autologous tumor lymphocyte; diagnosis; marker; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 61..342
FT /*tag= a
PE /product= "mamaglobin"
XX
WO9914230-Al.
XX
25-MAR-1999.
XX
18-SEP-1998; 98WO-US17991.
XX
18-SEP-1997; 97US-0933149.
XX
(UNIW ) UNIV WASHINGTON.
XX
Fleming TP, Watson MA;
XX
WPI: 1999-244021/20.
XX
P-PSDB; Y01718.
XX
Mamaglobin, secreted protein overexpressed in breast cancer
XX
Example 1; Fig 2; 60pp; English.
XX
The present sequence encodes a human mammary-specific protein,
designated mamaglobin. The specification describes a protein
comprising a mamaglobin antigen that is recognized by B and/or
Tc cells specific for the natural, secreted and glycosylated form
of mamaglobin polypeptide. This protein, or recombinant vectors
that express it, are used in vaccines for treating mamaglobin-
expressing cancers, specifically of the breast. Such cancers can
also be treated using autologous tumor lymphocytes activated
ex vivo with a mamaglobin antigen, then returned to the
patient. Expression of mamaglobin is elevated in 27% of stage I
primary breast cancers, so it represents a marker useful for
diagnosis of this disease.
XX
Sequence 503 BP; 146 A; 118 C; 97 G; 142 T; 0 other;
XX
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0
Percent Similarity: 84.932 Percent Identity: 53.425
XX
alignment_block:
US-09-367-009-3 x X26966
XX
Align seg 1/1 to: X26966 from: 1 to: 503
XX
2 SerGlyCysLysLeuGluAspMetValGluLysThrIleAsnSerAs 18
|||||
118 TCTGGCGCCCTTATTGGAGAAATGTGATTTCCAGAGAAATCAATCCACA 167

```

Copied from PCT009367009 on 27-04-2004

```

18 pIleSerIleProGluTyrLysGluLeuGlnGluPheIleAspSerA 35
|||||
168 ATGTGCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGAGACA 217
|||||
35 sPaLaAlaAlaGluAMetGlyLysPheLysGlnCysPheLeuAsnGln 51
|||||
218 ATGCCACTACAAATGCCATAGATGAATTCGAAGGAATGTTTCTTAACCAA 267
|||||
52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
|||||
268 ACGGATGAAGAACTCTGAGCAATGTTGAGGTGTTTATGCAATTAATATATGA 317
|||||
68 pSerIleTrpCysAsnLeu 74
|||||
318 CAGCACTCTTTGTGATTGA 336
seq_name: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT:AL2632
seq_documentation_block:
ID_AL2632 standard; CDNA; 503 BP.
XX
AC_AL2632;
XX
25-JUL-2000 (first entry)
XX
DE cDNA encoding the mammary-specific protein mamaglobin.
XX
KW Human; mamaglobin; mammary gland; breast cancer; endometrial cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 61..342
FT /*tag= a
PE /product= "mamaglobin"
XX
WO200018783-Al.
XX
06-APR-2000.
XX
29-SEP-1999; 99WO-US22616.
XX
29-SEP-1998; 98US-0162622.
XX
(UNIW ) UNIV WASHINGTON.
XX
Watson MA, Fleming TP;
XX
WPI: 2000-293105/25.
XX
P-PSDB; Y84622.
XX
Methods for detecting breast cancer, comprising detecting elevated
concentrations of a mamaglobin polypeptide, using an antibody, or
detecting elevated concentrations of the mRNA encoding the polypeptide,
using oligonucleotides -
XX
Example 1; Fig 2; 71pp; English.
XX
The present sequence encodes the human mammary-specific secreted
protein mamaglobin. Mamaglobin expression is restricted to the
mammary gland. Dysregulation of the mamaglobin gene occurs early
and frequently in breast cancer. The specification describes a method
for detecting the presence of breast cancer in a patient, comprising
detecting an elevated concentration of mRNA encoding a mamaglobin
polypeptide. The methods are useful for detecting the presence of
breast and endometrial cancer.
XX
Sequence 503 BP; 146 A; 118 C; 97 G; 142 T; 0 other;
XX
alignment_scores:
Quality: 208.00 Length: 73
Ratio: 3.355 Gaps: 0

```

Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:

US-09-367-009-3 x AL2632 ..

Align seg 1/1 to: AL2632 from: 1 to: 503

```
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
118 TCTGGCTGCCCTTATTGGAGAAATGTCATTCCAAAGACAATCAATCCACA 167

18 pIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAspSerA 35
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
168 AGTGTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCATAGACGACA 217

35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
218 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 267

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
268 ACGGATCAAACTCTGAGCAATGTTGAGGTGTTATGCAATTAAATATATGA 317

68 pSerIleTrpCysAsnLeu 74
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
318 CAGCAGTCTTTGTGATTGA 336
```

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:FL7694

documentation_block:

FL7694 standard; CDNA; 511 BP.

FL7694;

13-MAR-2001 (first entry)

Human breast cancer associated 19439-1 coding sequence.

Human; breast cancer associated gene; vaccine; diagnosis; therapy; 'ss.

Homo sapiens.

WO200060076-A2.

12-OCT-2000.

15-FEB-2000; 2000WO-US05308.

02-APR-1999; 99US-0285480.

23-JUN-1999; 99US-0339338.

02-SEP-1999; 99US-0389681.

03-NOV-1999; 99US-0433826.

(CORI-) CORIXA CORP.

Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;

WPI; 2001-122627/13.

An isolated polypeptide useful for the treatment and diagnosis of tumors e.g. breast cancer comprises at least an immunogenic portion of a breast tumor protein -

Claim 66; Page 146; 238pp; English.

The present invention provides the coding sequences and some protein sequences of proteins associated with breast cancer in humans. These sequences can be used in the diagnosis and treatment of cancers, particularly breast tumours.

Sequence 511 BP; 143 A; 98 C; 115 G; 155 T; 0 Other;

alignment_scores:

Quality: 208.00 Length: 73

Ratio: 3.355 Gaps: 0

Percent Similarity: 84.932 Percent Identity: 53.425

alignment_block:

US-09-367-009-3 x FL7694 ..

Align seg 1/1 to: FL7694 from: 1 to: 511

```
2 SerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSerAs 18
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
239 TCTGGCTGCCCTTATTGGAGAAATGTCATTCCAAAGACAATCAATCCACA 288

18 pIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAspSerA 35
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
289 AGTGTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTTCATAGACGACA 338

35 spAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsnGln 51
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
339 ATGCCACTACAAATGCCATAGATGAATGAAGGAATGTTTCTTAACCAA 388

52 SerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTyrAs 68
:|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
389 ACGGATCAAACTCTGAGCAATGTTGAGGTGTTATGCAATTAAATATATGA 438

68 pSerIleTrpCysAsnLeu 74
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
439 CAGCAGTCTTTGTGATTGA 457
```


gb_est87:BF417380	-	163.00	387.99	1.6e-12	469	!	BF417380	UI-R-CNO-bl-d-e-05-0
gb_est87:BF417713	-	163.00	387.97	1.6e-12	470	!	BF417713	UI-R-CNO-bl-r-c-07-0
gb_est87:BF417790	-	163.00	387.92	1.6e-12	472	!	BF417790	UI-R-CNO-bl-m-c-01-0
gb_est86:BF411207	-	163.00	387.75	1.7e-12	480	!	BF411207	UI-R-CNO-bl-m-d-05-0

```

seq_documentation_block:
LOCUS       BEO44893               351 bp      mRNA      08-JUN-2000
DEFINITION  hn08C04.x1 NCICAP.Thy6 Homo sapiens cDNA clone IMAGE:3021510.3'
             similar to TR:075556_075556 MAMMAGLOBIN B PRECURSOR. ; mRNA
             sequence.
ACCESSION   BEO44893
VERSION     BEO44893.1   GI:8361946
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 351)

```

REFERENCE	1 (cases 1 to 351)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike Emmert-Buck M.D. Ph.D. Vlad Knezevic M.D.

CDNA Library Preparation: Krizman Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

```

cDNA library Arrayed by: the I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 351.
Location/Qualifiers
1..351
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3021510"
/clone_lib="NCI CGAP Thy6"
/tissue_type="normal epithelium"
/lab_host="DH10B"
/note="Organ: Thyroid; Vector: pAMP10; mRNA made from
normal thyroid epithelium, cDNA made by oligo-dT priming.
Non-directionally cloned into UDG sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
cDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."
98 a 81 c 83 g 89 t

BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 390.00
  Ratio: 5.270
  Percent Similarity: 100.000
  Length: 74
  Gaps: 0
  Percent Identity: 98.649

alignment_block:
  US-09-357-009-3  x BE044893

```

```
alignment_block:
US-09-367-009-3 x BE044893
..
Align seq 1/1 to: BE044893 from: 1 to: 351
```

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
 |||||
 59 GATTCTGGTGTCAAACTCCTGGAGGACATGGTTGAAAAGACCATCAATTC 108
 17 rAspIleSerIleproGlnTvrIvrsGluLeuLeuGlncluphetIleAsps 34

```

|||||
109 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTTCATAGACA 158
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
159 GTGATGCCGTGCAGAGGCTATGGGAAATTCAGAGAGTTTCTCTCAAC 208
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
209 CAGTCACATAGAACTCGAAAACCTTGGACTGATGATGCATACAGTGTA 258
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
259 CGACAGCATTTGGTGTAAATAG 280
|||||
seq_name: gb_est69:BE044895

seq_documentation_block: 365 bp mRNA 08-JUN-2000
LOCUS BE044895
DEFINITION hn08403.x1 NCI-CGAP_Thy6 Homo sapiens cDNA clone IMAGE:3021509 3'
similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ;, mRNA
sequence.
ACCESSION BE044895
VERSION BE044895.1 GI:8361948
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike
Emmert-Buck M.D. Ph.D., Vlado Knezevic M.D.
CDNA Library Preparation: Krizman Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyt not found
Seq primer: -400P from Gibco
High quality sequence stop: 364.
FEATURES
Location/Qualifiers
1..365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3021509"
/clone_lib="NCI-CGAP_Thy6"
/tissue_type="normal epithelium"
/lab_host="DH10B"
/note="Organ: thyroid; Vector: pAMP10; mRNA made from
normal thyroid epithelium, cDNA made by oligo-dT priming.
Non-directionally cloned into UDG sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
cDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 103 a 85 c 85 g 92 t
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

```

alignment_block:
US-09-367-009-3 x BE044895

Align seg 1/1 to: BE044895 from: 1 to: 365

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
59 GATTCTGGCTGCAAACTCCTGGAGGACATGCTTGAAGAAGACCATCAATTC 108
|||||
17 rAspIleSerIleProGluTrpLysGluLeuLeuGlnGluPheIleAsps 34
|||||
109 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTTCATAGACA 158
|||||
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
159 GTGATGCCGTGCAGAGGCTATGGGAAATTCAGAGAGTTTCTCTCAAC 208
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
209 CAGTCACATAGAACTCGAAAACCTTGGACTGATGATGCATACAGTGTA 258
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
259 CGACAGCATTTGGTGTAAATAG 280
|||||

seq_name: gb_est5:AA297402

seq_documentation_block: 396 bp mRNA 18-APR-1997
LOCUS AA297402
DEFINITION EST112936 Endometrial tumor Homo sapiens cDNA 5' end similar to
similar to steroid-binding protein, C3 chain, prostate, mRNA
sequence.
ACCESSION AA297402
VERSION AA297402.1 GI:1949735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Pelligrino,S.M.,
Moreno-Palanco,R.F., McDonald,L.A., Nguyen,D.T., Saudek,D.M., Shirley,R.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other_ESTs: THCL168663
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

```

FEATURES
Source
Location/Qualifiers
1..396
/organism="Homo sapiens"
/db_xref="ATCC (inhost):138082"
/db_xref="taxon:9606"
/clone.lib="Endometrial tumor"
/sex="female"
/dev_stage="adult"
/Note="Organ: endometrium; Vector: pBluescript SK-;
Site_1: ECORI; Site_2: XhoI"
BASE COUNT 111 a 95 c 93 g 96 t 1 others
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x AA297402 ..
Align seg 1/1 to: AA297402 from: 1 to: 396
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
92 GATTCTGGCTGCAAACTCCTGGAGGACATGGTTGAAAAGACCATCAATTC 141
17 rAspIleSerIleProGluTyrLysGluLeuGluGlnGluPheIleAsps 34
142 CGACATATCTATACCTGAAATACAAAGAGCTTCTCAAGAGTTCATAGACA 191
34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
192 GTGATCCCGCTGCAGAGGCTATGGGAAATTCGAAGCAGTGTTCCTCAAC 241
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
242 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATGATGATGTA 291
67 rAspSerIleTrpCysAsnLeu 74
292 CGACACATTTGGTGAATATG 313
seq_name: gb_est7:AA398560

seq_documentation_block:
LOCUS AA398560 435 bp mRNA EST 16-MAY-1997
DEFINITION 373f02.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727995
3', similar to TR:G1199596 G1199596 MAMMAGLOBIN. ; mRNA sequence.
ACCESSION AA398560
VERSION AA398560.1 GI:2051669
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R., and Wilson, R.
WashU-Merck Est Project 1997
Contact: Wilison RK
Unpublished (1997)
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LILNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amersham.

```

Email: Robert_Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 887 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 401.

Location/Qualifiers
1. 435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:942754"
/clone_lib="NCI-CGAP_Thyl"
/tissue_type="thyroid"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive thyroid tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
122 a 86 c 94 g 133 t

BASE COUNT
ORIGIN
alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649
alignment_block:
US-09-367-009-3 x AA493295/rev ..

Align seg 1/1 to reverse of: AA493295 from: 1 to: 435
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
408 GATTCTGGCTGCAAACTCCTGGAGGACATGTTGAAAAGACCATCAATTC 359
17 rAspSerIleProGluTyLysGluLeuGluGlnGluPheIleAsps 34
358 CGACATATCTATACCTGAATACAAGAGCTTCTTCAAGAGTTTCATAGACA 309
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
308 GTGATCGCGTCGAGAGGCTATGGGAAATTCAGAGCTGTTCTCTCAAC 259
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
258 CAGTCACATAGAACTCTGAAAAAATTTGGAGTGTATGATGCATACAGTGTA 209
67 rAspSerIleTrpCysAsnLeu 74
208 CGACAGCATTTGGTGTATATG 187

seq_name: 'gb_est21:AI491987'
seq_documentation_block:
LOCUS AI491987 458 bp mRNA EST 12-MAY-1999
DEFINITION to07f11.x1 NCI-CGAP_U2 Homo sapiens cDNA clone IMAGE:2178381 3' similar to SW:MAMG_HUMAN Q13296 MAMMAGLOBIN PRECURSOR. [1];, mRNA sequence.
ACCESSION AT491987
VERSION AT491987.1 GI:4393001
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 458)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 577 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 399.
POLYA=No.

FEATURES
source

Location/Qualifiers
1. 458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2178381"
/clone_lib="NCI-CGAP_U2"
/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: Salt; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
132 a 93 c 101 g 132 t

BASE COUNT
ORIGIN
alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649
alignment_block:
US-09-367-009-3 x AI491987/rev ..
Align seg 1/1 to reverse of: AI491987 from: 1 to: 458

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
404 GATTCTGGCTGCAAACTCCTGGAGGACATGTTGAAAAGACCATCAATTC 355
17 rAspSerIleProGluTyLysGluLeuGluGlnGluPheIleAsps 34
354 CGACATATCTATACCTGAATACAAGAGCTTCTTCAAGAGTTTCATAGACA 305
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
304 GTGATCGCGTCGAGAGGCTATGGGAAATTCAGAGCTGTTCTCTCAAC 255
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValTy 67
254 CAGTCACATAGAACTCTGAAAAAATTTGGAGTGTATGATGCATACAGTGTA 205
67 rAspSerIleTrpCysAsnLeu 74
204 CGACAGCATTTGGTGTATATG 183

seq_name: 'gb_est23:AI659370'
seq_documentation_block:
LOCUS AI659370 460 bp mRNA EST 15-DEC-1999
DEFINITION tucic09.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2250736 3' similar to TR:O75556 O75556 MAMMAGLOBIN B PRECURSOR. ;, mRNA

```

sequence.
ACCESSION   AI659370
VERSION     AI659370.1  GI:4762940
KEYWORDS    EST.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 460)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 559 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 448.
FEATURES             source
     1..460
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone_image="IMAGE:2250736"
         /clone_lib="NCI-CCAP_Pr28"
         /sex="male"
         /dev_stage="adult"
         /lab_host="DH10B"
         /note="Organ: prostate; Vector: p7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CCAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 110192-1101959, and 121928-1220615)."
     132 a  94 c  101 g  133 t
BASE COUNT
ORIGIN
alignment_scores:
    Quality: 390.00      Length: 74
    Ratio: 5.270        Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 98.649
Alignment block:
US-09-367-009-3 x AI659370/rev ..
Align seg 1/1 to reverse of: AI659370 from: 1 to: 460
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
406 GATTCTGGCTGCAAACTCCTGGAGGACATGTTGAAAAGACCATCAATTC 357
|||||
17 rAspSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
356 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTCATAGACA 307
|||||
34 erAspAlaAlaGluAlaMeGlyLysPheLysGlnCysPheLeuAsn 50
|||||
306 GTGATGCCGCTGAGAGGCTATGGGAAATTCAGGAGTGTTCCTCAAC 257
|||||
51 GlnSerHisArgThrLeuLysAsnProGlyLeuMetMethIsthrValty 67
|||||

```

```

256 CAGTCACATAGAACTCTGAAAAAAGCTTTGGACTGATGATGATACATACGTGTA 207
67 rAspSerIleTrpCysAsnLeu 74
|||||
206 CGACAGCATTTGGTGTAAATATG 185
seq_name: gb_est6:AA393164
seq_documentation_block:
LOCUS      AA393164          491 bp      mRNA      EST      16-MAY-1997
DEFINITION zt73f02.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727995
5', similar to TR:G1199596 G1199596 MAMMAGLOBIN. ; mRNA sequence.
ACCESSION  AA393164
VERSION    AA393164.1  GI:2046134
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 491)
REFERENCE  1
AUTHORS   Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R., and Willson, R.
TITLE     WashU-Merck EST Project 1997
JOURNAL   Unpublished (1997)
COMMENT   Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     1..491
         /organism="Homo sapiens"
         /db_xref="GDB:5924907"
         /db_xref="taxon:9606"
         /clone_image="IMAGE:727995"
         /clone_lib="Soares_testis_NHT"
         /sex="male"
         /lab_host="DH10B"
         /note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGCCAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT  149 a  109 c  100 g  133 t
ORIGIN
alignment_scores:
    Quality: 390.00      Length: 74
    Ratio: 5.270        Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 98.649
Alignment block:
US-09-367-009-3 x AA393164 ..
Align seg 1/1 to: AA393164 from: 1 to: 491
1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
73 GATTCTGGCTGCAAACTCCTGGAGGACATGTTGAAAAGACCATCAATTC 122
|||||

```

Tue May 1 11:47:50 2001

```

17 rAspSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
123 CGACATATCTATACCTGATACAAAGAGCTTCTTCAAGAGTTCATAGACA 172
|||||
34 erAspAlaLaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
173 GTGATGCCGCTGCAGAGGCTATGGGAATTCAGCAGTGTTCCTCAAC 222
|||||
51 GlnSerHisArgThrLeuLysAsnGlyLeuMetMetHisThrValty 67
|||||
223 CAGTCACATAGAACTCTGAAAACTTTGGACTGATGATGATGATGATGTA 272
|||||
67 rAspSerIleTyrCysAsnLeu 74
|||||
273 CGACAGCATTTGGTGAATATG 294
|||||

seq_name: gb_est46:AW451131

seq_documentation_block: 494 bp mRNA EST 17-FEB-2000
LOCUS AW451131
DEFINITION UI-H-B13-alg-h-05-0-01.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2736992 3', mRNA sequence.
ACCESSION AW451131
VERSION AW451131.1 GI:6991907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 494)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI_CGAP clone distribution information can be found through the
I. M. A. G. E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=yes.

Location/Qualifiers
1..494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2736992"
/lab_host="NCI_CGAP_Sub5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clonides
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonides 1323912-1325891, 1471368-1472903
1492104-1493255); NCI_CGAP_Lus pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clonides 1414920-1417991, 1520904-1522439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Clonides 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonides 985608-986759
, 1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonides 1057416-1061255
, 1144584-1145351). (10% of the driver population), plus a

```

Plate: 229
Seq primer: Forward.
Location/Qualifiers
Source
1..636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGI"
/note="Vector: pBluescriptSKm"
BASE COUNT 189 a 152 c 134 g 161 t
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x AW966513 ..
Align seg 1/1 to: AW966513 from: 1 to: 636

```

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
91 GATCTGGCTGCAAACTCCTCGAGGACATGGTTGAAAAGACCAATCAATC 140

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
141 CGACATATCTATACCTGAATACAAAGAGCTTCTCAAGAGCTTCATAGACA 190

34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
191 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGACAGTGTTCCTCAAC 240

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
241 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 290

67 rAspSerIleTrpCysAsnLeu 74
|||||
291 CGACAGCATTTGGTGAATATG 312

```

seq_name: gb_est53:AW966509

seq_documentation_block: 647 bp mRNA EST 01-JUN-2000
LOCUS AW966509
DEFINITION EST378583 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.

ACCESSION AW966509
VERSION AW966509.1 GI:8156345
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 647)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 229
Seq primer: Forward.
Location/Qualifiers
Source
1..647
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGI"
/note="Vector: pBluescriptSKm"
BASE COUNT 193 a 155 c 140 g 159 t
ORIGIN

alignment_scores:
Quality: 390.00 Length: 74
Ratio: 5.270 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.649

alignment_block:
US-09-367-009-3 x AW966509 ..
Align seg 1/1 to: AW966509 from: 1 to: 647

```

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
91 GATCTGGCTGCAAACTCCTCGAGGACATGGTTGAAAAGACCAATCAATC 140

17 rAspIleSerIleProGluTyrLysGluLeuLeuGlnGluPheIleAsps 34
|||||
141 CGACATATCTATACCTGAATACAAAGAGCTTCTCAAGAGCTTCATAGACA 190

34 erAspAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
191 GTGATCGCGCTGCAGAGGCTATGGGAAATTCAGACAGTGTTCCTCAAC 240

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
241 CAGTCACATAGAACTCTGAAAACCTTTGGACTGATGATGATACAGTGTA 290

67 rAspSerIleTrpCysAsnLeu 74
|||||
291 CGACAGCATTTGGTGAATATG 312

```

seq_name: gb_est51:AW779377

seq_documentation_block: 458 bp mRNA EST 12-MAY-2000
LOCUS AW779377
DEFINITION hn79q08.x1 NCI CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3034142 3', similar to TR:O75556 O75556 MAMMAGLOBIN B PRECURSOR. ;, mRNA sequence.

ACCESSION AW779377
VERSION AW779377.1 GI:7793980
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 458)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 458.

FEATURES
Location/Qualifiers
1..458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3034142"


```

/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2459819"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      132 a      94 c      99 g      132 t      1 others
ORIGIN

alignment_scores:
  Quality: 385.00      Length: 74
  Ratio: 5.274        Gaps: 0
  Percent Similarity: 98.649      Percent Identity: 97.297

alignment_block:
  US-09-367-009-3 x AW779377/rev ..

Align seg 1/1 to reverse of: AW779377 from: 1 to: 458

1 AspSerGlyCysLysLeuLeuGluAspMetValGluLysThrIleAsnSe 17
|||||
404 GATTCTGGCTGCAAACTCTGAGGACATGTTGAAAGACCATCAATTC 355

17 rAspIleSerIleProGluTyrLysGluLeuGluInGluPheIleAsps 34
|||||
354 CGACATATCTATACCTGAATACAAAGAGCTTCTTCAAGAGTTATAGACA 305

34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
304 GTGATGCCGTGCAGAGGCTATGGGAAATTCGAAGCAGTGTTCCTCAAC 255

51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
254 CAGTCACATAGAACTCTGAAAACTTTGGAGTGTATGATGATGATGATGATG 205

67 rAspSerIleTrpCysAsnLeu 74
|||||
204 CGACAGCATTTGGTGTAAATAG 183

seq_name: gb_est27:AI936084

seq_documentation_block:
LOCUS      AI936084      471 bp      mRNA      02-SEP-1999
DEFINITION w61d06.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2459819 3'
            similar to TR:075556 O75556 MAMMAGLOBIN B PRECURSOR. ; mRNA
            sequence.
ACCESSION  AI936084
VERSION    AI936084.1 GI:5674954
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 471)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center

```

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 433.

FEATURES

source Location/Qualifiers

1..471

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="IMAGE:2459819"

/clone_lib="NCI_CGAP_Pr22"

/sex="male"

/tissue_type="normal prostate"

/lab_host="DH10B"

/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)

with a modified polylinker; 1st strand cDNA was prepared

from normal prostate bulk tissue, and was then primed with

a Not I - oligo(dT) primer. Double-stranded cDNA was

ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Not I and Eco RI sites of the

modified pT73 vector. Library is normalized, and was

constructed by Bento Soares and M. Fatima Bonaldo."

CAGCATTGGTGTAATAG 231

351 CGACATATCTATACC

Copied from PCT009367009 on 27-04-2004

Tue May 1 11:47:50 2001

```
34 erAspAlaAlaAlaGluAlaMetGlyLysPheLysGlnCysPheLeuAsn 50
|||||
301 GTGATGCCGCTGCAGAGCTATGGGAAATTCAAAGCAGTGTTCCTCAAC 252
|||||
51 GlnSerHisArgThrLeuLysAsnPheGlyLeuMetMetHisThrValty 67
|||||
251 CAGTCACATAGAACTCTGAAAAACTTTGGACTGATGATGCATACAGTGTA 202
|||||
67 rAspSerIleTrpCysAsnLeu 74
|||||
201 CGACAGCATTTGGTGTAAATATG 180
|||||
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 15:12:13 ; Search time 12.98 seconds
(without alignments)
109.523 Million cell updates/sec

Title: US-09-367-009-3
Perfect score: 392
Sequence: 1 DSGCKLMDVETKINSDIS.....TLKNFGLMHTVYDSIWCNL 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	99.5	95	3	US-08-821-451A-6
2	390	99.5	95	4	US-09-263-810-6
3	208	53.1	93	1	US-08-455-896-2
4	208	53.1	93	2	US-08-933-149-2
5	208	53.1	93	2	US-09-082-343-2
6	208	53.1	93	3	US-09-082-253-2
7	208	53.1	93	5	PCT-US96-08235-2
8	164	41.8	95	3	US-08-821-451A-27
9	164	41.8	95	4	US-09-263-810-27
10	163	41.6	95	2	US-08-455-896-7
11	163	41.6	95	2	US-08-933-149-7
12	163	41.6	95	2	US-09-082-343-7
13	163	41.6	95	3	US-09-082-253-7
14	163	41.6	95	5	PCT-US96-08235-7
15	61.5	15.7	1257	4	US-08-947-823-5
16	59.5	15.2	114	1	US-08-031-399-3
17	59.5	15.2	114	1	US-08-031-399-6
18	59.5	15.2	114	1	US-08-031-399-12
19	59.5	15.2	114	1	US-08-393-305-3
20	59.5	15.2	114	1	US-08-393-305-6
21	59.5	15.2	114	1	US-08-726-817-3
22	59.5	15.2	114	1	US-08-726-817-6
23	59.5	15.2	114	1	US-08-504-042-3
24	59.5	15.2	114	1	US-08-504-042-6
25	59.5	15.2	114	1	US-08-504-042-12
26	59.5	15.2	114	1	US-08-725-969-3
27	59.5	15.2	114	2	US-08-725-969-6

28	59.5	15.2	114	2	US-08-794-524-3	Sequence 3, Appl1
29	59.5	15.2	114	2	US-08-794-524-6	Sequence 6, Appl1
30	59.5	15.2	114	4	US-09-189-193-3	Sequence 3, Appl1
31	59.5	15.2	114	4	US-09-189-193-6	Sequence 6, Appl1
32	59.5	15.2	114	5	PCT-US94-03793-3	Sequence 3, Appl1
33	59.5	15.2	114	5	PCT-US94-03793-6	Sequence 6, Appl1
34	59.5	15.2	114	5	PCT-US94-03793-12	Sequence 12, Appl1
35	59.5	15.2	122	1	US-08-300-903A-3	Sequence 3, Appl1
36	59.5	15.2	122	1	US-08-031-399-2	Sequence 2, Appl1
37	59.5	15.2	162	1	US-08-031-399-5	Sequence 5, Appl1
38	59.5	15.2	162	1	US-08-393-305-2	Sequence 2, Appl1
39	59.5	15.2	162	1	US-08-393-305-5	Sequence 5, Appl1
40	59.5	15.2	162	1	US-08-535-733-2	Sequence 2, Appl1
41	59.5	15.2	162	1	US-08-284-393B-9	Sequence 9, Appl1
42	59.5	15.2	162	1	US-08-726-817-2	Sequence 2, Appl1
43	59.5	15.2	162	1	US-08-726-817-5	Sequence 5, Appl1
44	59.5	15.2	162	1	US-08-504-042-2	Sequence 2, Appl1
45	59.5	15.2	162	1	US-08-504-042-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-821-451A-6
; Sequence 6, Application US/08021451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESS: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,451A
FILING DATE: March 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,724
FILING DATE: March 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-821-451A-6

Query Match 99.5%; Score 390; DB 3; Length 95;
Best Local Similarity 98.6%; Pred. No. 5.3e-43;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DSGCKLEDVYKTSINSISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 60
 Db 19 DSGCKLEDVYKTSINSISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 78
 Oy 61 LMHTVYDSIWCNL 74
 Db 79 LMHTVYDSIWCNM 92

RESULT 2

US-09-263-810-6
 ; Sequence 6, Application US/09263810
 ; Patent No. 6174992

GENERAL INFORMATION:

APPLICANT: JIAN NI, Guo-Liang Yu and Reiner Gentz
 TITLE OF INVENTION: Human Endometrial Specific Steroid-
 TITLE OF INVENTION: Binding Factor I, II and III
 NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ADDRESSEE: CECCHI, STEWART & OLSTEIN
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/263,810
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/821,451
 FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.
 REGISTRATION NUMBER: 33,073
 REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 95 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-09-263-810-6

Query Match 99.5%; Score 390; DB 4; Length 95;
 Best Local Similarity 98.6%; Pred No. 5.3e-43;
 Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DSGCKLEDVYKTSINSISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 60
 Db 19 DSGCKLEDVYKTSINSISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 78
 Oy 61 LMHTVYDSIWCNL 74
 Db 79 LMHTVYDSIWCNM 92

RESULT 3

US-08-455-896-2
 ; Sequence 2, Application US/08455896
 ; Patent No. 5668267

GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.
 APPLICANT: FLEMING, TIMOTHY P.
 TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
 TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:

ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
 STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
 CITY: ST. LOUIS
 STATE: MISSOURI
 COUNTRY: USA
 ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/455,896
 FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 952726
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (314) 727-5188
 TELEFAX: (314) 727-6092

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 93 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-08-455-896-2

Query Match 53.1%; Score 208; DB 1; Length 93;
 Best Local Similarity 53.4%; Pred. No. 1.2e-19;
 Matches 39; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Oy 2 SGCCKLEDVYKTSINSISPEYKELQEFIDSDAAAEAMGKFKQCFLNOSHRTLNKFG 61
 Db 20 SGCPLNVISKTIPOVSKTEYKELQEFIDSDATTAIDELKECFUNQDETLSNVEV 79

Oy 62 MMHTVYDSIWCNL 74

Db 80 FMOLIYDSSLCDL 92

RESULT 4

US-08-933-149-2
 ; Sequence 2, Application US/08933149
 ; Patent No. 592836
 ; GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.
 APPLICANT: FLEMING, TIMOTHY P.
 TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
 TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:

ADDRESSEE: HOWELL & HAFERKAMP, L.C.
 STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
 CITY: ST. LOUIS
 STATE: MISSOURI
 COUNTRY: USA
 ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS